

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 15:03:05 ; Search time 3969 seconds
(without alignments)
11022.438 Million cell updates/sec

Title: US-09-758-269-5
Perfect score: 1800
Sequence: 1 atggctttttcacggcaac.....tggcgaagcagtggtga 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	589.8	32.8	837	28	BH62445
2	504	28.0	627	9	AV826228
3	459.6	25.5	766	29	BZ457573
4	425.6	23.6	720	12	BM412731

5	417	23.2	805	12	BM408615
6	416.2	23.1	781	28	BH49344
7	412.4	22.9	721	28	BH739063
8	393.8	21.9	644	12	BM535408
9	382.8	21.3	643	10	AW933524
10	375.6	20.9	787	12	BM408565
11	356.2	19.8	667	29	AL950790
12	344	19.1	592	10	BF113346
13	334.2	18.6	696	28	BZ022957
14	332.4	18.5	565	10	BE461924
15	326	18.1	637	13	BU550566
16	324.2	18.0	562	12	BM536135
17	323.2	18.0	553	12	BM085672
18	318	17.7	559	12	BM085005
19	316.6	17.6	520	10	BE432853
20	313.8	17.4	549	12	BM891057
21	312.8	17.4	742	28	BH559796
22	312.8	17.4	594	28	BH458011
23	309.4	17.2	547	10	BE437072
24	308.6	17.1	547	10	BE437072
25	304.8	16.9	578	29	EX004482
26	303.4	16.9	617	10	BE458861
27	300.4	16.7	546	12	BM084948
28	296.8	16.5	532	12	BJ563195
29	295.8	16.4	618	9	AW443298
30	295.8	16.4	711	29	BZ424431
31	294	16.3	564	12	BM085488
32	292.8	16.3	627	10	BF051297
33	289.2	16.1	509	10	BE451573
34	288.8	16.0	495	10	BE434930
35	285.2	15.8	649	29	CC016051
36	282.8	15.7	567	13	BO582721
37	281.4	15.6	651	29	BZ462969
38	277.8	15.4	641	29	BZ525251
39	276.2	15.3	490	10	BF050563
40	275.4	15.3	657	9	AA556214
41	267.4	14.9	503	28	BZ7476
42	267.4	14.9	602	13	BQ505126
43	265.8	14.8	668	29	CC157378
44	262.8	14.6	612	29	BZ525247
45	254.8	14.2	566	10	BE459895

ALIGNMENTS

RESULT 1
BH62445/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH62445 837 bp DNA linear GSS 19-FEB-2002
BOHTU09TR BO_2_3 KB Brassica oleracea genomic clone BOHTU09,
genomic survey sequence.

BH62445 GI:18721309

GSS.

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 837)

Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

FEATURES

```
source 1..837
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000D3"
/db_xref="taxon:3712"
/clone_lib="BOHTU09"
/notes="Vector: pROSL, Site 1: BstXI, 2-3 kb sheared genomic DNA inserted into pROSL using BstXI linkers"
BASE COUNT 171 a 238 c 220 g 208 t
ORIGIN

Query Match 32.8%; Score 589.8; DB 28; Length 837;
Best Local Similarity 83.2%; Pred. No. 8.6e-169;
Matches 596; Conservative 0; Mismatches 137; Indels 4; Gaps 2;

QY 807 AGTTCAGATCATCTCCATGAGATTTAAACACCGTTCGTCGTTTGTATGACGA 866
DB |||
DB 837 AGTCAGATCATCTCCAAAGGGAGACTGAAACCGTTGGCGTTACGATTTCAACGACAG 778
QY |||
DB |||
QY 867 ATTAGATCCCAATGATTGCCACCCGAAAGTCACCCGGATCC-GGTGAATCTTCG 925
DB |||
DB 777 TTTAGATCCCAATGATCGCCACCCGAAAGTCACCCGGATCCGCGGAGCTATTCG 718
QY |||
DB |||
QY 926 CTTTAAGCTACGACGTCGTTTCAAGCCTTACTTAAATACCTTCCGATTTCAACGGACG 985
DB |||
DB 717 CTCTAAGCTACGACGTCGTTTCAAGCCTTCTTAAAGTACTTCAAGTCTCAACGGACG 658
QY |||
DB |||
QY 986 GAATTAATACCGGACGTCGAGATTCAGTGTATGATCAGCAACGATGATGACGATTCG 1045
DB |||
DB 657 GGATTAATACCGGATGTCGAGATTCAGTGTATGATCAGCAACGATGATGACGATTCG 598
QY |||
DB |||
QY 1046 CGATTACAGAGAACTTCGTCGTCGATCTGATGACCAAGTCTGTTTCAAGTTCGCGGAGA 1105
DB |||
DB 597 CTATAACGAGAACTTCGTCGTCGATATACCGGACGCAAGTCTGTTCAAGTTCGCGGAGA 538
QY |||
DB |||
QY 1106 TGAATCCCGTGGTCTCGGTGGTTTACGACAGAAAGTTCGCAAGTTCGCGGATTT 1165
DB |||
DB 537 TGAATCCCGTGGTCTCGGTGGTTTACGACAGAAAGTTCGCAAGTTCGCGGATTT 478
QY |||
DB |||
QY 1166 TAGCAAAATACCGGACGTCGAGATTCAGTGTATGATCAGCAACGATGATGACGATTCCTTCT 1225
DB |||
DB 477 TAGCAAAATACCGGACGTCGAGATTCAGTGTATGATCAGCAACGATGATGACGATTCCTTCT 418
QY |||
DB |||
QY 1226 GCTTCATCTCTGGAACGCTTGGGAGAGCCGAGAAAGTCAAGTTCGTCGATGAGGT 1285
DB |||
DB 417 GCTTCATCTCTGGAACGCTTGGGAGAGCCGAGAAAGTCAAGTTCGTCGATGAGGT 358
QY |||
DB |||
QY 1286 CTTGTATGACTCCACAGACTCAATTTTCAACGAGTCTGACGAGAAATCTCAAGAGTGTC 1345
DB |||
DB 357 CTTGTATGACTCCACAGACTCAATTTTCAACGAGTCTGACGAGAAATCTCAAGAGTGTC 298
QY |||
DB |||
QY 1346 TGTCTGAATCCGCTGAATCTCAAAACCGTGAATCAACTCGCGTCCGATCATCTCCA 1405
DB |||
DB 297 TGTCTGAATCCGCTGAATCTCAAAACCGTGAATCAACTCGCGTCCGATCATCTCCA 238
QY |||
DB |||
QY 1406 ACCGAAGATCAACAGTCAACTCGAAGCAGGATGTTCAACAGAAACATGCTCGGCGGTA 1465
DB |||
DB 237 ACCGAAGATCAACAGTCAACTCGAAGCAGGATGTTCAACAGAAACATGCTCGGCGGTA 178
QY |||
DB |||
QY 1466 AATCCAAATTCGTTTACTTGGCTTTTACCGAGCCGTCGAGTAAAGTCTCAGGATTCGTA 1525
DB |||
DB 177 AGACCAAGTTTCGTTTACTTGGCTTTTACCGAGCCGTCGAGTAAAGTCTCAGGATTCGTA 118
QY |||
DB |||
QY 1526 AGTTGATCTCACTCTGGAAGATTTAAGAAACATCTTTACGCGGATACCGTTACGGAG 1585
DB |||
DB 117 AGTTGATCTCTTTTACCGGAGAGTCAAGAAAGTATCTGACGCGGTCGTTACGGT 58
QY |||
DB |||
QY 1586 GAGAGCCTCTGTTTCTCCCGGAGAAG- - -GAGGAGAGGAGAGACGAAGGATACATCC 1639
DB |||
DB 57 GAGAGCCTCTGTTTCTCCCGGAGAAG- - -GAGGAGAGGAGAGACGAAGGATACATCC 1639
```

```

424 GAGGTTTCTTGTACCGCAGAGACTACCCGTTTCTTAAACGGGTGATCTCTAGT 483
425 |
426 |
427 |
428 |
429 |
430 |
431 |
432 |
433 |
434 |
435 |
436 |
437 |
438 |
439 |
440 |
441 |
442 |
443 |
444 |
445 |
446 |
447 |
448 |
449 |
450 |
451 |
452 |
453 |
454 |
455 |
456 |
457 |
458 |
459 |
460 |
461 |
462 |
463 |
464 |
465 |
466 |
467 |
468 |
469 |
470 |
471 |
472 |
473 |
474 |
475 |
476 |
477 |
478 |
479 |
480 |
481 |
482 |
483 |
484 |
485 |
486 |
487 |
488 |
489 |
490 |
491 |
492 |
493 |
494 |
495 |
496 |
497 |
498 |
499 |
500 |
501 |
502 |
503 |
504 |
505 |
506 |
507 |
508 |
509 |
510 |
511 |
512 |
513 |
514 |
515 |
516 |
517 |
518 |
519 |
520 |
521 |
522 |
523 |
524 |
525 |
526 |
527 |
528 |
529 |
530 |
531 |
532 |
533 |
534 |
535 |
536 |
537 |
538 |
539 |
540 |
541 |
542 |
543 |
544 |
545 |
546 |
547 |
548 |
549 |
550 |
551 |
552 |
553 |
554 |
555 |
556 |
557 |
558 |
559 |
560 |
561 |
562 |
563 |
564 |
565 |
566 |
567 |
568 |
569 |
570 |
571 |
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |

```

```

BZ457573 766 bp DNA linear GSS 13-DEC-2002
BONGT80TF BO.1.6.2 KB tot Brassica oleracea genomic clone BONGT80,
genomic survey sequence.
ACCESSION BZ457573
VERSION BZ457573.1 GI:26737071
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 766)
Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BONGT80TF
Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..766
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BONGT80"
/clone_lib="BO.1.6.2_KB_tot"
/note="Vector: pHD1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHD1 using BstXI linkers"
BASE COUNT 177 a 203 c 203 g 183 t
ORIGIN
Query Match 25.5%; Score 459.6; DB 29; Length 766;
Best Local Similarity 76.7%; Pred. No. 6.1e-129;
Matches 575; Conservative 0; Mismatches 174; Indels 1; Gaps 1;
QY 529 GTTCACGGCGTCAATTCGAACACGGTTCAGTACGCTGCGGGTTTACTCAGACT 588
DB 751 GATCAGCGGTCCGGTTTAAATACGGTTCGGTTAGTTATTCGTGCGGTTACCGAACA 692
QY 589 A-ACGGTTTGTTCAGGAACGTCATNTGGGTGCGCGGTTTCCCAAGCCATCGGTGA 647
DB 691 ACCCGGTTTGTTCAGGACGAGAGTCGCGCGGTCGTGTTTCCCAAGCGATAGGAGA 632
QY 648 GTTTCACGCCACACCGGTATTGCCGACTCATGTTTCTACGCCAGCTGCGACCGG 707
DB 631 GTTTCACGGACACTTGGGGATCGCTAAGCTTAAGTATTAACGACACGTGGGCTTTCCG 572
QY 708 TATAGTCGACCGGCGACACCGGATAGCTAACCGCGGTTTGGTCTATTTCATATGG 767

```

```

571 TTTACTCGACCCGACCCCAAGGACTCGGTGTGGCTAATCCCGTTTAGTCTATTTCACGG 512
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |

```

RESULT 4

BM412731

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Seq primer: T3.

Location/Qualifiers

1..720

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

BM412731

EST587047

tomato breaker fruit Lycopersicon esculentum cDNA clone

CLEG60N24 5' end, mRNA sequence.

BM412731

GI:18264350

EST.

Lycopersicon esculentum (tomato)

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon

1 (bases 1 to 720)

J., Bougri, O., Kirkness, E., Uterback, T., Van Aken, S., Ronning

C.M., Fraser, C.M., Martin, G.B., Tankalev, S.D. and Giovannoni, J.

Generation of ESTs from tomato fruit tissue, breaker stage (2002)

Unpublished

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: http://www.genome.clemson.edu/orders/index.html

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

Location/Qualifiers

1..720

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEG50N24"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /clone_lib="tomato breaker fruit"
 /note="Vector: pBluescriptSKmCvAdapt; Site 1: EcoRI;
 Site 2: XhoI; supplier: Boyce Thompson Institute;
 sequencing: The Institute for Genomic Research. Fruit
 were harvested at the breaker stage (first sign of
 lycopene accumulation on the blossom end of fruit). Fruit
 were cut in half and the seeds and locules were discarded
 prior to freezing the pericarp."
 BASE COUNT 226 a 143 c 155 g 196 t
 ORIGIN

Query Match 23.6%; Score 425.6; DB 12; Length 720;
 Best Local Similarity 74.4%; Pred. No. 1.5e-118;
 Matches 536; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
 QY 740 CTAACGCCGCTTGGTCTATTTCATGGCGGTTATTGGCTATGTCGGAGGATGATTAC 799
 Db 1 CAACGCCGCTTGGTCTATTTCATGGCGGTTATTGGCTATGTCGGAGGATGATTAC 60
 QY 800 CTTACCAAGTTCAGATCACTCCCAATGGAGATTAAAAACCGTTGGTTCGATTTTG 859
 Db 61 CTTACCAATGTAAGGTAACACCCACCGCGGATCTTAAAAACAGAGGTCGATTCGATTTTCG 120
 QY 860 ATGGACAATTAGAAATCCACAATGATGCCACCCGGAAGTCCACCGGAATCCGTTGAAC 919
 Db 121 ACGCCAGCTAAATCCACCATGATAGTCCACCAAGCTCCACCCAGTTCCCGTGAGC 180
 QY 920 TCTTCGCTTAAAGTACGAGCTGCTTCAAGCTTTACCTAAATACCTTCGATTCGATTCAC 979
 Db 181 TATTGTCTTAGCTAGATGTGATTCAGAAAGCATACCTCAAGTACTTCAGATTTTCAA 240
 QY 980 CGGACGGAATTAATACCGGAGTCCAGATTCAGCTTCAGTCCAGCAAGCATGATGACG 1039
 Db 241 ABAATGGGAAATCAATGATGTGAATTCAGTTCAGACCCCAACATGATGATG 300
 QY 1040 ATTTCGCGATACAGAAATTCGTGCTGCTAGCTGACCAAGTCTGTTTCAAGCTGC 1099
 Db 301 ATTTCGCAATTAAGTACGAAATTCGTGCTGCTAGCTGACCAAGTCTGTTTCAAGATGT 360
 QY 1100 CGGAGATGATCCCGGCTGCTCCGCTGCTTACGCAAGCAAGGTCGCAAGATTCG 1159
 Db 361 CTGAATGATCCGTTGGAGGTTTCCCGGTTGTTTACGCAAGCAAGGTTCCCGATTTG 420
 QY 1160 GGAATTTAGCAAAATACCGCGAAGATTCATCGAAATTTAAGTGGATTTGATGCTCCAGATT 1219
 Db 421 GTATTCTGGATTAAGTACGCAAGATGGGTTCTGATTTGAATGGGTTGAAGTACCTGATT 480
 QY 1220 GCTTCTGCTTCCATCTCTGGAAGCTTGGGAAGAGCCAGAAACAGATGAAGTCGCTGA 1279
 Db 481 GTTCTGTTTCCACCTCTGGAATGCTTGGGAAGAGCAAGAAACAGATGAATCGTTGTA 540
 QY 1280 TAGGGTCTGTATGATCCTCAGATCAATTTTCAACGAGTCTGACGAGATCTTCAGA 1339
 Db 541 TTGTTTCATGATGACACCAAGATCCATTTTCAATGAATGTGATGAAGGCTTAAAGA 600
 QY 1340 GTGTCTGCTGTAATCCGCTGAATCTCAAAACCGGTGAATCAACTCCCGCTCCGATCA 1399
 Db 601 GTGTTTATCGAAATCCGCTCAATTTTGAACAGGGAATTCACAAGAAATCCATAA 660
 QY 1400 TCTCCAGGAGATCAACAGTCACTCAAGCAGGATGTTCAACAGAAATGCTGCTG 1459
 Db 661 TCGAAACCCCGGATGAACAAAGTGGATTTAGAAGCTGGAATGGTGAACCCGAAACAAACTCG 720

RESULT 5
 BM408615
 LOCUS EST582842 tomato breaker fruit Lycopersicon esculentum cDNA clone
 DEFINITION cLEG50N24 5' end, mRNA sequence.

ACCESSION BM408615
 VERSION BM408615.1
 KEYWORDS GI:18260245
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 REFERENCE 1 (bases 1 to 805)
 AUTHORS Alcala, J., Vrebalov, J., White, R., Vision, T., Karanymcheva, S.A., Tsai, J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Rønning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

TITLE Generation of ESTs from tomato fruit tissue, breaker stage (2002)
 JOURNAL Unpublished
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 This clone is available through the Clemson University Genomics Institute

Seq primer: T3.
 Location/Qualifiers
 1..805
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEG50N23"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /clone_lib="tomato breaker fruit"
 /note="Vector: pBluescriptSKmCvAdapt; Site 1: EcoRI;
 Site 2: XhoI; supplier: Boyce Thompson Institute;
 sequencing: The Institute for Genomic Research. Fruit
 were harvested at the breaker stage (first sign of
 lycopene accumulation on the blossom end of fruit). Fruit
 were cut in half and the seeds and locules were discarded
 prior to freezing the pericarp."
 BASE COUNT 262 a 144 c 180 g 219 t
 ORIGIN

FEATURES
 source
 1..805

Location/Qualifiers
 1..805
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEG50N23"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /clone_lib="tomato breaker fruit"
 /note="Vector: pBluescriptSKmCvAdapt; Site 1: EcoRI;
 Site 2: XhoI; supplier: Boyce Thompson Institute;
 sequencing: The Institute for Genomic Research. Fruit
 were harvested at the breaker stage (first sign of
 lycopene accumulation on the blossom end of fruit). Fruit
 were cut in half and the seeds and locules were discarded
 prior to freezing the pericarp."
 BASE COUNT 262 a 144 c 180 g 219 t
 ORIGIN

Query Match 23.2%; Score 417; DB 12; Length 805;
 Best Local Similarity 71.8%; Pred. No. 7.1e-116;
 Matches 575; Conservative 0; Mismatches 220; Indels 6; Gaps 2;
 QY 899 TCGACCCGGAATCCCGGTGAACCTCTTCGCTTAAAGCTACGAGCTGCTTCAAGCCTTACC 958
 Db 6 TCGACCCGATTTCCCGTGAGCTATTTCCTCTTAGCTAGCTGCTGATTTCAGAAGCCATACC 65
 QY 959 TAAATACTTCGGATTTCTCACCAGGAGAACTAAATCACCAGGAGCTCGAGATTGAGCTTG 1018
 Db 66 TCAAGTACTTCAGATTTTCAAAAATGGGAAATCAATGATGTTGAATTTCAAGTTG 125
 QY 1019 ATCAGCCAAAGATGATGACGATTTTCGGGATTTACAGAACTTCGTCTGCTGCTGCTGAC 1078
 Db 126 AAGACCCAAATGATGATGATGATTTTCGCAATTTACTGAGAACTTCGTCTGCTGCTGATC 185
 QY 1079 AGCAAGTCGTTTTTAAAGCTCGCGGAGATGATCGCGGTGGGCTCTCGGTGGTTTACGACA 1138
 Db 186 ACAAAGTCGTTTTCAAGATGCTGGAATGATCCGTTGAGGTTTCCCGTGGTTTACGACA 245
 QY 1139 AGAACAAGTCGCAAGATTCGGGATTTTAGCAAAATACGCCGAGATTCATCGAACATTA 1198
 Db 246 AGAACAAGATTTCCCGATTTGGTATTCTGGATAAGTACGCCGAAAGATGGGTCTGATTGA 305
 QY 1199 AGTGGATTCGATTCGAGATTCGCTTCGCTTCATCTCTGGAACCTTGGAAAGAGGCGAG 1258
 Db 306 AATGGGTTGAGTACCTGATTTCTGTTTCCACCTCTGGAATGCTTGGGAGAGAGCAG 365
 QY 1259 AACAAGATGAGTCGTCGATGATGAGGCTCTGATGACTCCACGAGACTCAATTTTCAAG 1318

Db 366 AACAGATGAATCGTTGTAATGGTTTCATGTATGACACACCAGACTCCATTTTCAATG 425
 Qy 1319 AGTCTGACGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTG 1378
 Db 426 AATGTGATGAGGCTAAAGAGTGTCTTATCCGAAATCCGCTCAATTTGAACACAGGA 485
 Qy 1379 AATCAATCGCGCTCGATCATCTCCACGAGAGTCAACAGTCAACCTCGAAGAGGGA 1438
 Db 486 AATCAACAAGAAATCCATAATCGAAACCCGGATGAACAGTGAATTTAGAGCTGAA 545
 Qy 1439 TGCTCAACAGAAATGCTCGCGGTAAACCAATTCGTTACTTTCGCTTTAGCCGAGC 1498
 Db 546 TGGTGAACCGGAAACAAATCGGAAGGAAACAGAGTATGCTTATTTGGCTATCGCTGAAC 605
 Qy 1499 CGTGCCTTAAAGTCTCAGGATTCGCTAAAGTGTATCTCACTACTGGAAGATTAAGAAAC 1558
 Db 606 CATGCCAAAAGTTCTCGTTTTCGAAAGTAAACCTGTTCAACCGGTGAAGTTGAGAAAT 665
 Qy 1559 ATCTTTACGGGATACCGTTACGGAGAGAGCTCTGTTCTCCCGGAGAGGAGGA - 1617
 Db 666 TCATTTATGGTGACAAACAAATATGTTGGGGAACCTCTTTTTTACAAGAGAACCAACAG 725
 Qy 1618 ----GAGGAACAGGAAGATACATCTCTGTTTCGTTCAACGAGAGAGACATGGAATC 1673
 Db 726 CCAGGAAGAGACGATGTTATTTTAGCTTTGTTTCAGCATGAGAA-AGATGGGAATC 784
 Qy 1674 GGAGTTACAGATAGTTAACGC 1694
 Db 785 AAAACTGCAATTTGGTAACGC 805

RESULT 6
 BH549344
 LOCUS BOGRQ53TR BOGR Brassica oleracea genomic clone BOGRQ53, genomic survey sequence.
 ACCESSION BH549344
 VERSION BH549344.1 GI:17801124
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 781)
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished
 COMMENT Other_GSSs: BOGRQ53TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..781
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGRQ53"
 /clone_lib="BOGR"
 /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
 BASE COUNT 204 a 203 c 192 g 182 t

Query Match 23.1%; Score 416.2; DB 28; Length 781;
 Best Local Similarity 70.8%; Pred. No. 1.2e-115;

Matches 553; Conservative 0; Mismatches 228; Indels 0; Gaps 0;
 Qy 274 GGGCGGGCGGAGCGTTTGACGCGCGGAGGGTTTCTTGTTCAGCCAGAGACTACAC 333
 Db 1 GCAGCGCGGATTCGATCGATGTGCTGAGCGAGGTTGATCTCAGCTGAGCAATACT 60
 Qy 334 CGCTTCTTAAACCGGCTGATCTAGTGTTCAGATCGCGGAAATTTTGTCCGCTGAAT 393
 Db 61 CTTCTTCCAAACCGCTGATCCAGTATTCAAATCGCGGGAATTTATTTCCCGGTGGC 120
 Qy 394 GAACAGCCGTCGCGGTAATCTTCGGTGTGCGGAAACTTCCCGATTCATCAAGGA 453
 Db 121 GAATTTCCGTCGAGAGTGCTTCGACGTTGAAGGAACAATCCCTGACTGCAATTAACGA 180
 Qy 454 GTGTATGTGCGCAACGAGCTTAACCCACTTCACGAGCGGTGACAGGTCACACTTCTTC 513
 Db 181 GTTATGTCCGTAACCGGTGCAATCCGATGTTTCGAGCCAAATCGCTGGGCATCATTTGTT 240
 Qy 514 GACGAGAGCGGTATGTTTCAACCGCTCAATTCGAACACGGTTCAGCTAGCTACGCTTGC 573
 Db 241 GACGAGAGCGGAATGTTTCACGCACTTAAATAACCAACGGTTCAGCAAGCTACGATGC 300
 Qy 574 CGCTTCTTCACTAACCGGTTTTCAGGAACGTCATTTGGGTGACCGGTTTTCCTCC 633
 Db 301 CGCTTCTTCACTAACCGGAGATTAATTTCAAGAAACAGATTGGGTGACCGGTTTTCG 360
 Qy 634 AAAGCATCGGTGAGCTTCACGCGCACACCGGTATTGCCGACTCATGCTATTTACGCG 693
 Db 361 AAAGCAATCGGAGGCTTCACGCGCACTCGGGAATCGCACGTCGTATGTTTTTACGCA 420
 Qy 694 AGAGCTGACGCGGTATAGTCGACCGGCAACCGGTAACCGGTGTAGCTAACCGCGTTTG 753
 Db 421 CGTGGCTTTTGGCGCTCGTCAACAATCAAAACGGCTCGGGTAGCGAACCGCGTTTG 480
 Qy 754 GTCATTTTCAATGCGCGTTATGCTATGTCGAGGATGATTTACCTTACCAAGTTCAG 813
 Db 481 GTTACTTCAATAACCGGCTTTTAGCAATGTGAGAAAGAGATTTACCGTACCAATTA 540
 Qy 814 ATCACTCCCAATGGAGATTTAAACACCGTTGGTTCGGTTCGATTTTATGAGCAATTA 873
 Db 541 ATAACCTAAACCGGCGACCTCCAAACCGTTGGCGGTACGATTTCCGCGTCAGTTAAA 600
 Qy 874 TCACAAATGATTCGCCACCGAAGTCGACCGGATTCGGGTGAATCTTCGCTTTAAGC 933
 Db 601 TCATCATGATCGCCACCGAATTCGACCGGTTTACAAAGAGCTCCACCGCTTAAC 660
 Qy 934 TAGCAGCTCGTTTCAAAGCTTACCTTAAATACTTCCGATTTCTCACCGACGGAATA 993
 Db 561 TAGCAGCTCGTTTAAAGGCTTACCTGAATATTTTCAAGATTTCTCGCGGCGGTTAAA 720
 Qy 994 TCACCGAGCTCGAGATTCAGCTTGTATGATGACCAACGATGATGACGATTTCCGATT 1053
 Db 721 TCACCGAGCTCGAGATTCGCGCTCGAGACTCCGACGATGTTTCCAGATTTCCGTTAACG 780
 Qy 1054 G 1054
 Db 781 G 781

RESULT 7
 BH739063
 LOCUS BOMHT72TF BO_2_3_KB Brassica oleracea genomic clone BOMHT72, genomic survey sequence.
 DEFINITION BOMHT72TF BO_2_3_KB Brassica oleracea genomic clone BOMHT72, genomic survey sequence.
 ACCESSION BH739063
 VERSION BH739063.1 GI:18844458
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 721)

Qy Dp

920 TCTTCGCTTTAAAGCTACGACGTCGTTCCAGAGCCCTTACCTAAATATACCTCCGATTCCTC
181 TATTTGCTCTTAGCTACGATGTGATTCCAGAAGCCATACCTCAAGTACTTCAGATTTTCAA 240

```

FEATURES
    source
        1. 643
            /organism="Lycopersicon esculentum"
            /mol_type="rRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="CLSP54E2"
            /tissue_type="fruit pericarp"
            /dev_stage="mature green (3-5 days pre-ripening)"
            /lab_host="SOUR"
            /clone_lib="tomato fruit mature green, TAMU"
            /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
            XhoI; cLEF - Fruit were tagged at the 1cm stage and
            harvested 3-5 days prior to ripening. Fruit were cut in
            half to verify the seeds were indeed 'immature' and the
            seeds and locules were discarded prior to freezing the
            pericarp"

```

CLEGA5E23 5' end, mRNA sequence.
ACCESSION
VERSION BM408565
KEYWORDS 1 GI:18260195
EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 787)
AUTHORS Alcalá, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai,
J., Bougri, O., Kirkness, E., Uterback, T., Van Xien, S., Renning,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage (2002)
JOURNAL Unpublished
COMMENT Contact: CUGI

Contact: CUGI
Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics Institute
Seq primer: T3.
Location/Qualifiers
1..787
/organism="Lycopersicon esculentum"
/mol_type="rRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG45E23"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/note="Vector: pBluescriptSKCudapt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."
BASE COUNT 216 a 194 c 165 g 212 t
ORIGIN
Query Match 20.9%; Score 375.6; DB 12; Length 787;
Best Local Similarity 67.8%; Pred. No. 3.3e-103;
Matches 525; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
QY 205 TCTCCGCGCATTTGTTTAAAGCCCAAGCAAGAAATCCACACATTAACAGATGATTTG 264
Db 5 TCACACCCAAAACAGAAACACACACTCTCTCTTCTTCACTTCCAAAGTGAATTA 64
QY 265 TTCACAGAGCGCGCGGAGGTTGGAGCGGGGAGGTTCTTGTTCAGCCACGAG 324
Db 65 GTCCAGAAAGCAGCAGCATGGTTAGATGCTGTAGAAGTCTTTAACTAAACATGA 124
QY 325 AAGCTACACCGGTTCTCTAAACGGGTGATCTCTAGTGTTCAGATCGCCGGAATTTGCT 384
Db 125 CTTGAACACCTTTGCGGAAACAGCCGACCCACGAGTCCAGATTTCTGGGAATTTGCT 184
QY 385 CCGGTGATGAACAGCCCGTCCGCGTATCTTCGGTGTGCGGAAACTTCCCGATTC 444
Db 185 CCGGTACCGGAAATTCAGTCTGTCAATCTCTCCGTCACCGGAAATATCCAAATGT 244
QY 445 ATCAAAGGAGTATGTGCGCAACGAGCTAACCCACTTCACGAGCCGTTGACAGTAC 504
Db 245 GTTCAAGCGGTTTACGTTGCAACGAGGCTAACCTCTTTTGAACCAACCGCGGACAC 304
QY 505 CACTTCTTCACGAGAGCGGTATGTTTCAGCCGTCATAATTCGAACACGTTTCAGTAGC 564
Db 305 CATTCTTCACGCGGACGGTATGGTTACGCGGTTCAATTTCAAAATGGTGGCTAGT 364
QY 565 TAGCTTTCGCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCATTTGGTTCGACCG 624
Db 365 TAGCTTTCGCGTTTCACTGAACAGAGAGCTTTGTTCAAGAAAAGCTTTGGTCCGCT 424
QY 625 GTTTTCCCAAGCATCGGTAGCTTCACGGCCACACCGGTATTCGCGACTCATGCTA 584
Db 425 GTTTTCCCTAAAGCCATTGGTGAATATCATGGTCACTCTGGAATTCAGGCTTATGCTG 484
QY 685 TTCTACGCCAGAGTGCAGCGGTTATAGTCGACCCGCGCACACGGAACCGGTGTAGCTAAC 744
Db 485 TTTTACGCTGTGGGCTCTTCGGACTTGTGTGATCACAGTAAGGAAGTGTGTGCAAC 544
QY 745 GC CGGTTTGGTCTAATTTCAATGGCCGGTTATTTGGCTATGTTCGAGGATGATTTACCTTAC 804
Db 545 GC CGGTTTGGTCTAATTTCAATTAACCGGATTTACTTGTATGTCTGAAGATGATTTGCCTTAC 604
QY 805 CAAGTTCAGATCACTCCCAATGGAGATTTAAACACCGTTGCTCGGTTTCGATTTGATGA 864

Db 605 CATGTAAGGTAACACCCACCGCGATCTTAAACAGAGGGTCTGATTCGATTCGACGCG 664
QY 865 CAATTAGATTCACATGATTCGCCACCGAAGTCGACCGGAATCCGGTGAATCTTTC 924
Db 665 CAGCTAAATCACCATGATAGCTACCCCAAGCTCGACCGGTTCCGGTAGTATTT 724
QY 925 GCTTTAAGCTAGCAGCTCGTTTCAAGCGCTTACCTAAATATCTCCGATTCCTCA 978
Db 725 GCTCTTAGCTAGCATGTGATTGAGAAGCCATACCTCCAGTACTTCAGATTTTCA 778
RESULT 11
AL950790 667 bp DNA linear GSS 24-OCT-2002
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-331D05-016046,
DEFINITION genomic survey sequence.
ACCESSION AL950790.1 GI:24407412
VERSION Arabidopsis thaliana (thale cress)
KEYWORDS Arabidopsis thaliana
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eutosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.
and Weissshaar, B.
A pipeline for automated high-throughput generation of FSTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
Unpublished
2 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
Unpublished
3 (bases 1 to 667)
Rosso, M., Strizhov, N., Li, Y. and Weissshaar, B.
Direct Submission
Submitted (21-OCT-2002) Weissshaar B., Max-Planck-Institut fuer
Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion close to or within gene Atg10100. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.
FEATURES
1..667
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-331D05-016046"
/note="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector pAC161. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"
BASE COUNT 176 a 145 c 161 g 184 t 1 others
ORIGIN
Query Match 19.8%; Score 356.2; DB 29; Length 667;
Best Local Similarity 71.3%; Pred. No. 2.6e-97;
Matches 469; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 509 TCTTCGAGGAGCGGTATGTTTCCGCGCTCAAAATCGAACACGGTTCAGTACG 569

```

Db      5  TATTCGACGGTGACGGTATGGTTTACGCGGTAAAAATCACTAAACGGAGACCGAGTTACT 64
QY      569 CTTCCGGGTTTACTCAGACTACCGGTTTGTTCAGAACGTCATATGGTTCGACGGGTTT 628
Db      65  COTCCGGGTTTACGGAACCGGAGATGGTTTCAAGAGAAACAATCGGTTCTCCGATTT 124
QY      629 TCCCAAAAGCCATCGGTGAGCTTTCACGGCCACACCGGTATTGCCGACTCATGTATTTCT 688
Db      125 TCCCTAAAGCTATAGGTGAGCTACATGTCATCTGGAATCGACGGATGATGCTATTTT 184
QY      689 AGCCAGAGCTCAGCCGGTATAGTCGACCCGGGACACGGACCGGTGTAGCTAACCGCG 748
Db      185 AGCCAGCGGTATTTTCGGTTTATTAATAATCAAAAACCGGAACCGGATTCCTAACCGCG 244
QY      749 GTTTCGTTCTATTTCAATGCCCGGTTTATGGCTATGTCGGAGGATGATTACTTACCAAG 808
Db      245 GTTTCGTTTACTTCCACGACCGGTTTATAGCTATGTCGTAAGATGATCTACTTACCAAG 304
QY      809 TTCAGATCACTCCCAATGAGATTTAAAAACGGTTGGTGGGTCGTTGATGATGACAAAT 868
Db      305 TTCGTGTCACTGACATGCGGATTTAGAGACCATCGGAAGATTCGATTTCGACGACAAC 364
QY      869 TAGAATCCACAATGATTGCCACCCGAAAGTCGACCCGGAATCCGGTGAACTCTTCGGTT 928
Db      365 TAAGCTCCGCAATGATCGCTACCCGAGATTTGATCCGTTACGAGAGATTTTGGGT 424
QY      929 TAAGCTACGATCGTTTCAAGCGCTTACCTAAATATCTCCGATTCATCCGACGGAA 988
Db      425 TGAGCTACGACGCTGTTAAGAAACCGTATTTGAAATATCTTTAAATTCCTCGCGGAAGTG 484
QY      989 CTAATACCCGACGCTCGAGATTCAGTTGATCAGCCACGATGATGATGATGATGATGATG 1048
Db      485 AGAATACCCGACGCTGAGATTCCTCGCCAGTCGACGATGATGATGATGATGATGATGATG 544
QY      1049 TTACAGAGAACTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1108
Db      545 TCACCTGAGAATTCGTTGTGATTCGGATCAACAAAGTTGTGTTTAAAGCTCCGATATGT 604
QY      1109 TCCGGGTCGGTCTCCGGTGGTTTACGACAGAAAGTCGACAGATTCGGGATTTT 1166
Db      605 TTCTTGGGAAATCTCCGGTTAAATACGACGGAGAGAAATTTCCCGGTTTGGATGTT 662

```

```

RESULT 12
LOCUS   Bf113346
DEFINITION Bf113346 tomato breaker fruit Lycopersicon esculentum cDNA clone
          CLE344B9 5' sequence, mRNA sequence.
ACCESSION Bf113346
VERSION   Bf113346.1 GI:10943036
KEYWORDS EST.
SOURCE   Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE 1 (bases 1 to 592)
AUTHORS  Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
          Liang, P., Hansen, T.S., Craven, M.B., Bowman, C.L., Renning, C.M.,
          Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley,
          S.D.
TITLE     Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL   Unpublished
COMMENT   Contact: CUGI
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall,
          Clemson, SC 29634, USA
          Email: http://www.genome.clemson.edu/orders/index.html.
          Location/Qualifiers
            1..592
              /organism="Lycopersicon esculentum"
              /mol_type="mRNA"

```

```

RESULT 13
BZ022957/c
LOCUS    oeh32c12.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BZ022957
VERSION   BZ022957.1 GI:23582691
KEYWORDS GSS.
SOURCE    Brassica oleracea
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

BASE COUNT 194 a 106 c 131 g 161 t
ORIGIN
Query Match 19.1%; Score 344; DB 10; Length 592;
Best Local Similarity 73.8%; Pred.No.1.3e-93;
Matches 437; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 893 CGAAAGTCACCCGGGAATCCGGTGAACCTCTTCGCTTTAAGCTACGACGTCGTTCAAAGC 952
Db 1 CAAAGCTCGACCCAGTTTCCGGTGAGCTATTTCCTTAGCTACGATGATTCAGAACG 60
QY 953 CTTACCTAAATACCTTCGATTCCTACCGGCGGAACTAAATCAACGGACGTCGAGATTC 1012
Db 61 CATACCTCAAGTACTTCAGATTTTCAAAAATGGGGAATAATCAATGATGTGAATTC 120
QY 1013 AGTTGATCAGCCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1072
Db 121 CAGTTGAAGACCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 1073 CTACACCAAGTCGTTTCAAGCTCCCGAGATGATGATGATGATGATGATGATGATGATGATG 1132
Db 181 CTGATCAACAAGTCGTTTCAAGATCTCTGAAATGATGATGATGATGATGATGATGATGATG 240
QY 1133 ACCACAAGAACAAAGTCGCAAGATTCGGATTTTAGACAAATACGCGCAAGATTCATCGA 1192
Db 241 ACACAGAACAAAGTTTCCGATTTGGTATTCGTAAGTACGCGAAGATGCGGTCTG 300
QY 1193 ACATTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1252
Db 301 ATTGAATGGGTGAAGTACCTGATGTTTCTGTTTCCACCTCTGGAATGCTTGGGAAG 360
QY 1253 AGCCAGAAACAGATGAAGTCGTCGTCATAGGGTCTGTATGATCTCCACAGACTCAATTT 1312
Db 361 AAGCAGAAACAGATGAATTCGTTGATTTGGTTCATGATGATGATGATGATGATGATGATGATG 420
QY 1313 TCAACGAGTCTGACGAGAACTCTCAAGAGTGTCTGTCTGAAATCCGCTGAAATCTCAAAA 1372
Db 421 TCAATGAATGTGATGAAGGCTAAAGAGTGTGTTTATCCGAAATCCGCTCAATTTGAAA 480
QY 1373 CCGGTGAATCAACTCCCGCTCGATCATCTCCAACGAAGATCAACAAAGTCAACCTCGAAG 1432
Db 481 CAGGGAATCAACAGAAATCCATTAATCGAAACCCGGATGACAAAGTGAATTTAGAAG 540
QY 1433 CAGGGAATGTCACAGAAACATGCTCGCGGTAAACCAAAATTCGTTTACTT 1484
Db 541 CTCGAATGGTGAACCCGAAACAAACTCGGAAGGAAACACAGATGATGCTTATTT 592

```

```

RESULT 13
BZ022957/c
LOCUS    oeh32c12.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BZ022957
VERSION   BZ022957.1 GI:23582691
KEYWORDS GSS.
SOURCE    Brassica oleracea
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 696)
REFERENCE
AUTHORS Delaunay, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash
W., Rabinowicz, P.D. and Wilson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oeh32 row: c column: 12
Seq primer: -28RPpOT reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 551.

Location/Qualifiers
1..696
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea 10100PH3 buds provided by
Thomas Osborn at the University of Wisconsin. genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
150 a 205 c 161 g 180 t

BASE COUNT
ORIGIN
Query Match 18.6%; Score 334.2; DB 28; Length 696;
Best Local Similarity 71.2%; Pred. No. 1.5e-90;
Matches 441; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
QY 793 GATTTCCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAACCGTTGTCGGTTC 852
DB 696 GATTTCGGGTATCAGCTGAATAATCAACACCGGGAATCTCGAACCGTTGGCGTTAC 637
QY 853 GATTTTGATGACAAATAGAAATCCAAATGATTCGCCACCGGAAATCGACCGGAATCC 912
DB 636 GATTTTCACGGTCACTTAAATATCCATGATCGCCACCGGAAATCGACCGGTTAGC 577
QY 913 GGTGAATCTTCGTTTAACTAGCTAGCGTCTTCAAGCTTACCTTAAATCTTCGGA 972
DB 576 AAGGAGCTCCACGGCTAAGCTAGCTCGTGAGAAAGCTTACTTGAAGTACTTCAGA 517
QY 973 TTCTCACCGGACGGAATAATACCGGACGTCGAGATTGAGTTGATCAGCCAAAGTG 1032
DB 516 TTCTCGCGGAGCGCGTTAAATACCGGAACTAGAGATCCCGTTCGAGACTCCGAGATG 457
QY 1033 ATGCACGATTCGGGATTACAGAACTTCGTGCTGCTACTGACGACGAGTCTTTTC 1092
DB 456 ATTACGATTTTCGTATTAACGGAATCTTCGTGTTGATTCGGATCAACAGTCGTGTT 397
QY 1093 AAGCTGCGGAGATGATCCGCGGTGGTCTTCGGTGTGTTTACGACAGAAACAAGTTCG 1152
DB 396 AAGTAGGGAGATGATCGCGGAACTCGCGGTGTTTTCGACGCGGATAGGTTTCG 337
QY 1153 AGATTGGGATTTAGACAAATACCGGAGATTTCATCGAACATTAAGTGGATTGCT 1212
DB 336 CGATTGGGATTAATGCCGGAAGACGCGACGAGGCTTCGAGATAATCTCGGTGGATT 277
QY 1213 CCAGATTCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGACGAGAAACAGATGAAGTC 1272
DB 276 CCGGAGAGTTCTGTTTCCATCTCTGGAACGCTTGGGAAGACGAGAGTGAAGTG 217
QY 1273 GTCGTGATAGGGTCTGATGACTCCACGAGACTCAATTTTCAACGAGTCTGACGAGAT 1332
DB 216 TTGTTGATCGGTGCTGATGTCGCGCGGATTTCTATCTTCAACGAGAGACGAGAGC 157
QY 1333 CTCAGAGTGTCTCTGAAATCGCGCTGAATCTCAAAACCGGTGAATCAACTCGCCGT 1392

Db 156 TTGAAAGCGTCTTTGACGGAGATCAGATAAACCTCAGGACGCTGAATCCACGCCCGA 97
QY 1393 CCATCATCTCCAAACGAAG 1411
Db 96 GCGATGATGTCGACGATG 78

RESULT 14
BE461924
LOCUS
DEFINITION BE461924 566 bp mRNA linear EST 18-MAY-2001
EST143439 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLEG40L2, mRNA sequence.
ACCESSION BE461924
VERSION BE461924.1 GI:9506322
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 566)
REFERENCE
AUTHORS Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Renning, C.M.,
Niernman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley
S.D.

Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

Location/Qualifiers
1..566
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG40L2"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit, TIGR"
/notes="Vector: pBluescriptSKmCtAdapt; Site 1: EcoRI;
Site 2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end
of the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
184 a 104 c 124 g 154 t

BASE COUNT
ORIGIN
Query Match 18.5%; Score 332.4; DB 10; Length 566;
Best Local Similarity 74.2%; Pred. No. 4.6e-90;
Matches 420; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 893 CGAAAGTCGACCCGGAATCCGGTGAACTCTTCGCTTTAAGCTAGACGTCGTTTCAAGC 952
DB 1 CAAGCTCGACCCAGTTTCCGGTGAGCTATTTCCTTAGCTACGATGTGATTGAGAGC 60
QY 953 CTTACTTAATATCTTCGCTGATTCACCGGACGAACTAAATCACCAGCTCGAGATTC 1012
DB 61 CATACCTCAAGTACTTTCAGATTTTCAAAATGGGAAATCAATGATGTTGAAATTC 120
QY 1013 AGCTTGATCAGCAACGATGATGACGATTCGGGATTCAGAGAACTTCGTCGTCGTAC 1072
DB 121 CAGTTGGAAGACCAACAATGATGATGATGATTTTCGCAATTTACTGAGAACTTCGTCGTCATTC 180
QY 1073 CTCGACGACGATCGCTTTTCAAGCTCCGAGATGATCCGCGGTGGGTCTCCGGTGGTTT 1132
DB 181 CTGATCAACAGTCGCTTTTCAAGATGTCGAAATGATTCGAGGTTTCAACGCGTGGTTT 240

```

QY 1133 ACGACAGACAAAGTTCGAGATTCGGGATTTAGACAAATACGCCGAGATTCATCGA 1192
Db 241 ACGACAGACAAAGTTCGAGATTTGGTATTCGGATAAGTAGCGGAAAGATGGGTCTG 300
QY 1193 ACATTAAAGTGATGATGCTCCAGATTCCTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252
Db 301 ATTTGAAATGGTGAAGTACCTGATTTCTGTTTCCACCTCTGGAATGCTTGGGAAG 360
QY 1253 AGCCAGAACAGATGAAGTCTGCTGATAGGCTCTGTATGACTCCACGAGATCAATTT 1312
Db 361 AAGCAGAACAGATGAAGTCTGTAATGCTTGTATGATGACACCAAGATCCATTT 420
QY 1313 TCACAGATCTGACGAGAAATCTCAAGAGTGTCTCTGTAATCCGCTCGAATCTCAAAA 1372
Db 421 TCATGATGTGATGAAGGCTTAAGAGTGTCTTATCCGAAATCCGCTCAATTTGAAA 480
QY 1373 CCGTGTATCAATCGCGCTCGATCATCTCCACGAGAGATCAACAGTCAACTCGAAG 1432
Db 481 CAGGAAATCAACAGAAATCCATATCGAAACCCGGATGAACAAAGTGAATTTGAAG 540
QY 1433 CAGGATGCTCAACAGAAACATGCTC 1458
Db 541 CTGGAATGCTCAACCGAAACAACTC 566

RESULT 15
BU550566/c
LOCUS
DEFINITION GM880021A20F06 Gm-r1088 Glycine max cDNA clone Gm-r1088-7932 3',
mRNA sequence.
ACCESSION BU550566
VERSION BU550566.1 GI:22933427
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 637)
Vodkin,L., Shoemaker,R., Keim,P., Retzel,E., Khanna,A., Shealy,R.,
Clough,S., Thibaud-Nissen,P., Coryell,V., Erpelding,J., Raph,C.,
Shoop,E., Stromvik,M., Schweitzer,P., Gong,G. and Liu,L.
A Functional Genomics Program for Soybean (NSF 9872565) (2002)
Unpublished
Other ESTs: BR99117 corresponding to Gm-cl068-3283 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
Insert Length: 637 Std Error: 0.00
Plate: GM880021A20 row: F column: 06
Seq primer: 5'-TTTTTTTCTTTTCTTTT(A/C/G)-3'
High quality sequence stop: 637.
Location/Qualifiers
1..637
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1088-7932"
/clone_lib="Gm-r1088"
/notes="The library Gm-r1088 is a sequence-driven, reracked
set of 9,216 low redundancy clones selected from cDNA
libraries from various tissues and stages of development
of soybean that consists of 2,706 cDNAs from germinating
cotyledons (source library Gm-cl027); 1,355 cDNAs from
immature seed coats (libraries Gm-cl019 and Gm-cl023); 917
cDNAs from tissue culture derived somatic embryos (source

```

libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from germinating seedlings, shoot tips, or leaves exposed to various stresses (source libraries Gm-cl064, Gm-cl065, Gm-cl066, and Gm-cl067); and Gm-cl068); and 963 cDNAs from young leaves exposed to bacterial and fungal pathogens (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rackted to form library Gm-r1088 and the cDNA clones of the reracked Gm-r1088 library were then sequenced at the 3' end. The unigene selection and 3' sequencing was funded by NSF Plant Genome project #9872565 (http://soybean.genomics.croscpi.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://129.186.26.94/soybeanest.html. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://web.ahc.umn.edu/biodata/nsfsoy/. Reracking and 3' sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

```

BASE COUNT 160 a 180 c 126 g 171 t
ORIGIN
Query Match 18.1%; Score 326; DB 13; Length 637;
Best Local Similarity 72.1%; Pred. No. 4.5e-88;
Matches 454; Conservative 0; Mismatches 170; Indels 6; Gaps 2;
QY 1140 GAACAGGTGCGCAAGATTCGGGATTTTGAACAATACGCCGAGATTCATCGAAATTA 1199
Db 637 GAACAGGTTCACAGATTCGGGATTCGACAAAGATGCTAAAGATGCAATATGATGAA 578
QY 1200 GTGGATTGATGCTCCAGATTCGTTCTGTTCCATCTCTGGACGCTTGGGAGAGCCAGA 1259
Db 577 ATGGATCGATGCTCCGAGTGTCTTGTTCCATCTCTGGACGCTTGGGAGAGCCAGA 518
QY 1260 AACAGATGAAGTCGTGATAGGTCCTGTATGATCAACAGACTCAATTTTCAACA 1319
Db 517 AATGATGAGATCGTTGTCTCGTATGATGATGATGATGATGATGATGATGATGATGAT 458
QY 1320 GTCTGACGAGATCTCAAGATGTCCTGTCTGAATCCGCTGAATCTCAAAACCGGTGA 1379
Db 457 ATCGAAGAGAGTTTGAAGAGCATCTTTCGGAGATAGGCTGAATTTGAAGACAGGCA 398
QY 1380 ATCAACTCCGCTCCGATCATCTCCAAAGAGATCAACAAGTCAACCTCGAAGCAGGAT 1439
Db 397 GTCCACGAGAAACCCATATCTC---GGAATCCGAAACAAGTCAACTTGAAGCCGGAT 341
QY 1440 GGTCAACAGAAACATGCTCGGCGTAAACCAATTCGCTTACTTGGCTTTAGCCGAGCC 1499
Db 340 GGTCAACAGAAACATGCTCGGAGAAAGACAGAGTTTCGGGTACTTAGCTTTCGGAGCC 281
QY 1500 GTGGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAAGTTAAGAAACA 1559
Db 280 GTGGCTTAAAGTTTCGGGTTTTCGCAAGTTTGTATTTGTTTCAGTGGGAGGTTTAAAGA 221
QY 1560 TCTTTACGGCGATAACCGTTAGCGAGAGAGCTCTGTTTCTCCC---CGGAGAAAGAGG 1616
Db 220 CATGTATGCGAAGAGAGGTTTCGGAGGGAGGCTCTGTTTCTTCCAAATGGTGTGACGG 161
QY 1617 AGAGGAAGACGAGAGGATACATCTCTGTTTCTGTTTACGACGAGAGACATGAAATTCGGA 1676

```

Db 160 TGATGAAGACGATGGCCATATTCTCGCATTCGTCATGATGAGAAAGAGTGGAAATCGGA 101
Qy 1677 GTTACAGATAGTTAACCGGTTAGCTTAGAGTTGAGCAACGGTTAAACTTCGGTCAAG 1736
Db 100 GCTGCAGATTGTCAATGCCAAACTTTGAAGCTTGAGGCTTCAGTTAGCTTCCTTCCAG 41
Qy 1737 GGTTCGTTACGGATTTTCACGGTACATTCAT 1766
Db 40 AGTTCCTTATGGGTTTCACGGTACTTTAT 11

Search completed: November 16, 2003, 18:30:16
Job time : 3974 secs

1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGAAACGCTTGGAGAGCCAGAA 1260
 Db
 1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGAAACGCTTGGAGAGCCAGAA 1260
 Qy
 1261 ACGATGAAGTGTGCTGATAGGCTCTGATGATGCTCCACAGAGCTCAATTTTCAACGAG 1320
 Db
 1261 ACGATGAAGTGTGCTGATAGGCTCTGATGATGCTCCACAGAGCTCAATTTTCAACGAG 1320
 Qy
 1321 TCTGACGAGAATCTCAAGAGTGTCTGCTGCTGAAATCCGCTGAAATCTCAAAACCCGGTGAA 1380
 Db
 1321 TCTGACGAGAATCTCAAGAGTGTCTGCTGCTGAAATCCGCTGAAATCTCAAAACCCGGTGAA 1380
 Qy
 1381 TCAACTCGCGCTCCGATCATCTCCACGAAGATCAACAGTCAACCTCGAAGAGGATG 1440
 Db
 1381 TCAACTCGCGCTCCGATCATCTCCACGAAGATCAACAGTCAACCTCGAAGAGGATG 1440
 Qy
 1441 GTCAACAGAAACATGCTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCG 1500
 Db
 1441 GTCAACAGAAACATGCTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCG 1500
 Qy
 1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTGTGATCTCACTACTGGAGAGTTAAGAAACAT 1560
 Db
 1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTGTGATCTCACTACTGGAGAGTTAAGAAACAT 1560
 Qy
 1561 CTTTACGCGATTAACCGCTTACGAGGAGAGCCCTCTCTTTCTCCCGGAGAGGAGGAG 1620
 Db
 1561 CTTTACGCGATTAACCGCTTACGAGGAGAGCCCTCTCTTTCTCCCGGAGAGGAGGAG 1620
 Qy
 1621 GAAGACGAAGATACATCTCTGCTTTGCTTTCACGAGAGAGATCGGAAATCGGAGTTA 1680
 Db
 1621 GAAGACGAAGATACATCTCTGCTTTGCTTTCACGAGAGAGATCGGAAATCGGAGTTA 1680
 Qy
 1681 CAGATAGTTAACGCGCTTACGCTTAGAGTTGAAGCAACGCTTAACTTCCGTCGAAGGTT 1740
 Db
 1681 CAGATAGTTAACGCGCTTACGCTTAGAGTTGAAGCAACGCTTAACTTCCGTCGAAGGTT 1740
 Qy
 1741 CCGTACGAGTTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCT 1800
 Db
 1741 CCGTACGAGTTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCT 1800

RESULT 2

US-09-758-269-15
 ; Sequence 15, Application US/09758269
 ; Patent No. US20020104120A1
 ; GENERAL INFORMATION:
 ; APPLICANT: IUCHI, SATOSHI
 ; APPLICANT: KOBAYASHI, MASATOMO
 ; APPLICANT: SHINOZAKI, KAZUO
 ; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
 ; FILE REFERENCE: 3914-3
 ; CURRENT APPLICATION NUMBER: US/09/758,269
 ; PRIOR FILING DATE: 2001-01-12
 ; PRIOR APPLICATION NUMBER: JP 2001-003476
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: JP 2000-010056
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 1818
 ; TYPE: DNA
 ; ORGANISM: Lycopersicon esculentum
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1815)
 US-09-758-269-15

Query Match 45.8%; Score 824.6; DB 10; Length 1818;
 Best Local Similarity 69.6%; Pred. No. 1.3e-273;
 Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;

121 GCCAGTGTGTACACGTAAGTCAATGTTTCACTGCGCTTCACTCTCCAGCTCTT 180
 Db
 121 GCCAGTGTGTACACGTAAGTCAATGTTTCACTGCGCTTCACTCTCCAGCTCTT 180
 Qy
 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCAATGTTTAAAGCCCAAGCAAGAA 240
 Db
 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCAATGTTTAAAGCCCAAGCAAGAA 240
 Qy
 241 TCAACACTAACAAGATGATGTTTCCAGAGAGCGGCGGAGCGTTCGAGCGCGCG 300
 Db
 241 TCAACACTAACAAGATGATGTTTCCAGAGAGCGGCGGAGCGTTCGAGCGCGCG 300
 Qy
 301 GAGGTTTCTCTGTGACGACACAGAGTACACCCGCTTCTTAAACCGCTCATCTAGT 360
 Db
 301 GAGGTTTCTCTGTGACGACACAGAGTACACCCGCTTCTTAAACCGCTCATCTAGT 360
 Qy
 361 GTTCAGATCGCGGAAATTTGCTCGGTGATGATGAACACCCGCTCGGCGTAACTTCCG 420
 Db
 361 GTTCAGATCGCGGAAATTTGCTCGGTGATGATGAACACCCGCTCGGCGTAACTTCCG 420
 Qy
 421 GTGGTCGGAACCTTCCGATTCATCAAGAGTGTATGTGCGCAACGAGCTTAACCCA 480
 Db
 421 GTGGTCGGAACCTTCCGATTCATCAAGAGTGTATGTGCGCAACGAGCTTAACCCA 480
 Qy
 481 CTTTACGAGCCGCTGACAGTCACTCTCTCGACGAGAGCGGTATGTTTCAACCGCTC 540
 Db
 481 CTTTACGAGCCGCTGACAGTCACTCTCTCGACGAGAGCGGTATGTTTCAACCGCTC 540
 Qy
 541 AAATTCGAAACAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 600
 Db
 541 AAATTCGAAACAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 600
 Qy
 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGAGCCATCGGTGAGCTTCAACGCCAC 660
 Db
 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGAGCCATCGGTGAGCTTCAACGCCAC 660
 Qy
 661 ACCGTTATCCCGACTGATGCTATTTACGCGCAGAGCTGACGCGGTATGTCGACCGG 720
 Db
 661 ACCGTTATCCCGACTGATGCTATTTACGCGCAGAGCTGACGCGGTATGTCGACCGG 720
 Qy
 721 GCACACGGAACCGGTGTAGTAAACCGGTTTGGTCTATTTCAATGCGCGGTTATGGCT 780
 Db
 721 GCACACGGAACCGGTGTAGTAAACCGGTTTGGTCTATTTCAATGCGCGGTTATGGCT 780
 Qy
 781 ATGTCGAGGATGATTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 840
 Db
 781 ATGTCGAGGATGATTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 840
 Qy
 841 GTTGTGCGTTCGATTTTGTGAGCAATTAGAATCCCAATGATTTGCCACCGGAAAGTC 900
 Db
 841 GTTGTGCGTTCGATTTTGTGAGCAATTAGAATCCCAATGATTTGCCACCGGAAAGTC 900
 Qy
 901 GACCCGGAATCCGCTGAACTCTTCCGTTTAAAGCTACGACGTCGTTTCAAGCCCTTACCTA 960
 Db
 901 GACCCGGAATCCGCTGAACTCTTCCGTTTAAAGCTACGACGTCGTTTCAAGCCCTTACCTA 960
 Qy
 961 AAATACCTTCCGATTTCTACCGGAGCGGAACTAAATCAGCGAGCTGAGATTCAGCTTGT 1020
 Db
 961 AAATACCTTCCGATTTCTACCGGAGCGGAACTAAATCAGCGAGCTGAGATTCAGCTTGT 1020
 Qy
 1021 CAGCCACGATGATGACGATTTCCGATTTACAGAGAACTTCTGCTGCTGCTGCTGCTGCT 1080
 Db
 1021 CAGCCACGATGATGACGATTTCCGATTTACAGAGAACTTCTGCTGCTGCTGCTGCTGCT 1080
 Qy
 1081 CAGCTCGTTTTCAGCTTCCGAGATGATCCGCGGTGGTCTCCGCTGGTTTACGACAAG 1140
 Db
 1081 CAGCTCGTTTTCAGCTTCCGAGATGATCCGCGGTGGTCTCCGCTGGTTTACGACAAG 1140
 Qy
 1141 AACAGGTCGAAGATTCGGATTTTACGAAAATACGCGGAGATTCATCGAATTAAG 1200
 Db
 1141 AACAGGTCGAAGATTCGGATTTTACGAAAATACGCGGAGATTCATCGAATTAAG 1200

170 CTCAGCTCTTCATTCCCTAAGCAATCATCAAACTCTCCGGCAATGTTGTTAAAGCCCA 229
185 CTTCAAAATTAACAACACCAAGAAATATACAAATTCACACCCCAACAAGAAACAACA 244
230 AAGCAAAGAAATCCAAACATAAACAGATGAATTTGTTCCAGAGAGCGGCGCGAGCGT 289
245 ACTCCTCTTCTTCTTCACTTCCAAAGTGAATTTAGTGCAGAAAGCAGCAATGGCTT 304
290 TGGACGGGCGGAGGTTTCTCTGTACGACACGAGAGCTACACCGCTTCTTAAACGG 349
305 TAGATGCTGTAGAAGTCTTTAACTAAACATGAATTTGAACACCTTTCGCGAAACAG 364
350 CTGATCTCTAGTGTTCAGATCCCGGAAATTTGTCGGGTGAATGAACAGCGCTCCGGC 409
365 CCGACCCACGAGTCCAGATTTCTGGAAATTTGTCGGGTACCGGAAATCCAGTCTGTG 424
410 GTAATCTTCCGGTGTGGGAAATCTCCGATTCATCAAGAGAGTGTATGTGCGCAACG 469
425 AATCTTCTCCGGTCAACGGAATAATACCCAAATGTGTCAAGGGTTTACGTTGGAACG 484
470 GAGCTAACCCACTTCAAGAGCGGTGACAGTCAACCACTTCTTCGAGGAGACGGTATGG 529
485 GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCAATTTCTCGACGGCGACGGTATGG 544
530 TTCAGCGCGTCAAAATTCGAAACCGGTTTCAAGTGTGCTGCTTACTCAGACTA 589
545 TTCAGCGCGTCAAAATTCGAAACCGGTTTCAAGTGTGCTGCTTACTCAGACTA 604
590 ACCGCTTTGTTTCAGGAACGTCATTTGGGTTCGACCGGTTTCCCAAGGCCATCGTGAGC 649
605 AGAGCTTTGTTCAAGAAAGCTTTGGGTGCGCTGTTTCCCTAAGGCCATTTGGTGAAT 664
650 TTCAGCGCACACCGGTATGCGCACTATGCTATTTCTAGCCAGAGCTCGACCGGTA 709
665 TACATGGTCACTCTGGAATTCGAGGCTTATGCTGTTTACGCTCGTGGGCTTTCGAC 724
710 TAGTCGACCGGCAACGAGCGGTAGCTTAACCGCGGTTTGGTCTTATTTCAATGGCC 769
725 TTGTTGATCACAGTAAGAACTGGTGTGCAACGCGGTTTACTCTATTTCAATACC 784
770 GGTATTTGCTATGTCGAGGAGTATTTACCTTACCAAGTTTCAATCTCCCAATGAG 829
785 GATTAATCTGCTGCTGAAGATGATTTGCGCTTACCATGTAAAGTAAACCCACCGCGC 844
830 AATTAAACCGGTGCTGGTTCGATTTGATGGAATTAAGATCCCAATGATTTGCC 889
845 ATCTTAAACAGAGGTGCAATTCGATTTGACGCGGACGCTAAATCCACCATGATGTC 904
890 ACCGAAAGTCGACCGGAAATCGGTGAATCTTTCGCTTTAAGTACGAGCTGCTTCAA 949
905 ACCGAAAGTCGACCGGAAATTCGCTGAGCTATTTGCTTACGATGATGATTCAGA 964
950 AGCTTACCTTAAATACCTTCCGATTTCTACCGGAGGAACTAAATCACCGGACGTCAGA 1009
965 AGCATACCTTCAAGTACTTCAATTTCAAAATGCGGAAATCAATGATGTGAA 1024
1010 TTCAGCTTATGATCAGCAACGATGATGACGATTTCCGATTTACAGAACTTCGTCGTCG 1069
1025 TTCAGTTGAAGACCCCAACATGATGCAATGATTTGCAATTTACTGAGAACTTCGTCGTC 1084
1070 TACTCAGCAGAGTCTGTTTCAAGCTCCGAGATGATCCGCGGTGCTCCGCTGG 1129
1085 TTCTGATCAACAGTCTGTTTCAAGATGCTGAAATGATCCGCTGAGGTTTCCCGGTGG 1144
1130 TTTTACGACAGAAACAAAGTCTGGAATTTCCGATTTTGAACAAATACCGCAAGATTCAT 1189
1145 TTTTACGACAGAAACAAAGTCTGGAATTTCCGATTTTGAATTTGATGATGCGAAAGTGGT 1204
1190 CGAACTAATAGTGGATGATGCTCCAGATGCTTCTGCTTCCATCTCTGAGACGCTGGG 1249
1205 CTGATTTGAATGGGTTGAAGTACCTGATTTCTGTTTCCACCTCTCGAATGCTGGG 1264
1250 AAGAGCCAGAAACAGATGAAGTCTGCTGATAGGCTCTGATGAGTCTCCACGAGACTCAA 1309

1265 AAGAAGCAGAAACAGATGAATCGTTGTAATGGTTTCATGTATGACACACGACTCCA 1324
1310 TTTTCAACGAGTCTCAGAGAAATCTCAAGAGTGTCTCTCTGAAATCGGCTGAAATCTCA 1369
1325 TTTTCAATGAATGATGATGAAGGCTAAAGAGTGTGTTATCCGAATCGTCTCAATTTGA 1384
1370 AAACCGGTGAATCAACTCGCGTCCGATCATCTCAAGAGATCAACAGTCAACCTCG 1429
1385 AAACGAGGAATCAACAAGAAATCCATAATCGAAACCCGATGAACAGTGAATTTAG 1444
1430 AAGCAGGATGCTCAACAGAAACATGCTCGGCGTAAACCAAAATCGCTTACTTTGGCTT 1489
1445 AAGCTGGAATGTGAAACCGAAACAACTCGAGAGAAACAGAGTATGCTTATTTGGCTA 1504
1490 TAGCCGAGCGGTGGCTTAAAGTCTCAGATTCGCTTAAAGTTCATCTACTCTGAGAG 1549
1505 TCGCTGAACCATGCGCAAAAGTTTCTGGTTTTTGAAGAGTAAACCTGTTCACCGGTGAAG 1564
1550 TTAAGAAACATCTTACGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGCGAG 1609
1565 TTGAGAAATTCATTTATGTGTACAAATATGTTGGGGAACCTCTTTTACCAGAG 1624
1610 -----AAGGAGGAGAGGAAAGCAAGGATACATCTCTGTTTCTGTCACGACGAGAGA 1663
1625 ACCCAACAGCAAGGAAGAGCAGATGTTATTTTAGCTTTCGTTACGATGAGAAAG 1684
1664 CATGGAATCGGATTTACAGATAGTTAAACCGGTTAGCTTAGAGTTGAAGCAACGGTTA 1723
1685 AATGGAATCAGAACTGCAAAATGTTTACGCAATGAGTTTGAAGTTGAGGCAACTGTA 1744
1724 AACTTCGCTCAAGGTTCCGTTACGAGATTTCAAGGATACATCTGTTTCTGTCACGACGAG 1783
1745 AGCTTCAACAGAGTTCTTATGATTTTCAAGCAATTCATTAACGCAATGATTTGG 1804
1784 CGAAGCAGG 1792
1805 CAAATCAGG 1813

RESULT 3

US-09-758-269-11
; Sequence 11, Application US/09758269
; Patent No. US2002104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZIO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1839
; TYPE: DNA
; ORGANISM: Vigna unguiculata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1836)
US-09-758-269-11

Query Match 44.4%; Score 798.4; DB 10; Length 1839;
Best Local Similarity 70.4%; Pred. No. 1.5e-264;
Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;
QY 233 CCAAAGAATCCACACATAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGGCGTGG 292

281 CCRACCAACATTAACCTCAAAATGGAACTTTCTCAGAAAGCGCTGCGCGCTTGG 340
293 ACGCGGCGAGGTTTCTTGTGACGACGAGAGCTACACCGGCTTCTAAACCGGTG 352
341 ACCTGGTCGAAACCGGCTCTGTCTGACGAGCGCAACACCGCTCCCAAAACCGCGG 400
353 ATCTAGTGTTCAGATCGCGGAAATTTTCTCGGTGAATGAACAGCCCGTCCGCGTA 412
401 ACCGAGGTCCTCAATCGCCGGAATCTTCGCGCGGTGCGGAGCATGCGCGATCAAG 460
413 ATCTTCGCTGTCGGAATCTTCGCGATTCATCAAGAGTGTATGTGCGCAACGAG 472
461 GACTCCCGGTGTCGGAATATCCCAATTCGATTCGACGCGGTGTACGTGCGCAACG 520
473 CTAACCCACTTCACGACCGGTGACAGCTCACCACTTCTTCGACGAGAGCGGTATGGTTC 532
521 CCAATCCGCTCTACGAGCTGTGCGCGGCAACCACTTCTTCGAGCGGACGCGATGGTCC 580
533 ACGCGTCAATTCGAACAGCGTTTACGTAGCTACGTTTCCGCGTTTACTCAGATAACC 592
591 ACGCGGTGAAGTTCAGAACCGGCGCGCCAGCTACGCTTCCGCTTTCACCGAGAGCGAGC 640
593 GTTGTGTTCAAGAACGTCATAGGTCGACCGGTTTCCCAAGCCATCGGTGAGCTTC 652
641 GTCTCTCGAGGAGAACTCTAGGCGCGCGGTTCGCGAGGCCATCGGCGAGCTCC 700
653 ACGGCCACCGGTATTCGCGACTCATGTATTTACGCGCAGAGCTGACGCGGTATAG 712
701 ACGGCCACTCGGATCGCGCGCTCCTCTCTTACGCGCGCGGTCTTTCGCGGCTCG 760
713 TCGACCGCGCACACGAGCGGTGTAGCTAACCGCGGTTTGGTCTATTTCATATGCGCGGT 772
761 TTGATGGTCCAGGGCATGGGCGTGGCGAACCGCGGTCTCGTCTACTTCAACAACCAACC 820
773 TATTGGCTATGTCGAGAGATGATTACCTTACCAAGTTCAGATCATCTCCCAATGAGATT 832
821 TCITGGCGCATGTCGAGAGAGATTTACCTTACCACGTGGAATCACCCCTTAACCGCGACT 880
833 TAAACACCGTGTGCGGTTCGATTTTGATGGACAAATAGAAATCCCAATGATTCGCCACC 892
881 TAACCAACGTTGGCGGTACGACTTCAACGGGCGAGTCACTCAACAATGATCGCCACC 940
893 CGAAGTCGACCGGATCGGTGAATCTTCGCTTTAAGCTACGACGTCGTTTCAAGC 952
941 CGAATCGGACCGGTCGACGGGACCTCCACCGCTCAGCTACGACGCTATTCAGAGC 1000
953 CTTACCTAAATATCTTCGATTTCTCACCGGACGGAATCAATCACCGGACGTCGAGATTC 1012
1001 CTTACCTCAAGTACTTCGTTTCTCCCGCGGCGTCAAGTCCCGCGAGTGGAAATCC 1060
1013 AGCTTGATCAGCAACGATGATGACGATTTTCGGATTACAGAACTTCGTCGTGATC 1072
1061 CCCTGAAGGACCCACCATGATGACGATTTGCGCATACGGAGAAATTCGTCGTGCTCC 1120
1073 CTGACCAAGAGTCTGTTTCAAGTCGCGGAGATGATCCGCGTGGGTCTCCGCGTGGTTT 1132
1121 CCGACCAAGAGTGGTCTTCAACTCAACGAGATGATCACCGCGGTCCTCCCGTGGTCT 1180
1133 ACGCAGAGAAAGGTGCAAGATTCGGGATTTAGCAATACGCGGAGATTCATCGA 1192
1181 ACGCAGAGAAACAACTTCAGGTTTGGGATTTGCGATTCGCAAGAAATGCGAGGACGGAATG 1240
1193 ACATTAAGTGTATGATGCTCCAGATTCGTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252
1241 CGATGCGGTGATCGACGCGCGGATTTTCTGCTTCCACCTCTCGAAGCGGTGGAGG 1300
1253 AGCAGAGAAAGATGATGCTGATAGGTCCTGTATGATCTCCACAGATCATTT 1312
1301 AGCCGGAACCGAGAGGTTGTGTGATGGTCTCTCATGACCCCTTGGGACTCCATTT 1360
1313 TCAACGAGTCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCGTGAATCTCAAA 1372

1361 TCAACGAATGCGAGGAGGTTTGAAGAGCGTCTGTCTCAGAGATAAGGCTGAACCTTGAGGA 1420
1373 CCGGTGAATCAACTCGCGTCCGATCATCTCCAAAGAGATCAACAAGTCAACCTCGAAG 1432
1421 CCGGCAAGTCCACTCGGCGCCCATTTCTCGAGCGCG--AAAGTGAACCTGGGAG 1477
1433 CAGGATGTCTAACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492
1478 CCGCATGTGTGAACAGAAACAGTTCGGAAGAGACCCAGTTCGCTATCTGGCTCTGG 1537
1493 CCGAGCGGTGGCTTAAGTCTCAGGATTCGCTTAAAGTTGATCTACTACTCGAGAGTTA 1552
1538 CCGAGCGCTTGGCCCAAGTCTCGGCTTTGCGAAAGTTGATTTGCTAGTGGGAGTGA 1597
1553 AGAAACATCTTTACGCGGATAACCGTTACGAGGAGAGCCTCTGTTTCTCCCGGAGAG 1612
1598 AGAAGTACATGTATGAGAGAGAAAGTTCGCTGGGAGGCTCTGTTTCTTCTCC--AACG 1654
1613 GAGGAGAGAGAGAGAGGATCATCTCTGTTTCTGTTTCAAGAGAGAGACATGGAAAT 1672
1655 GCCAAAGAGAGAGATGGGTATTTCTGGCANTCTGTCACGACGAGAAAGATGGAAAT 1714
1673 CGGATTTACAGATAGTTAAACCGCTTAGCTTAGAGGTTGAAGCAACGTTAAACTTCGT 1732
1715 CCGAGCTGAGATTTGTAATGCCAAATTTAAAGCTGAAAGCTTCCATCAACTCCCT 1774
1733 CAAGGCTTCGCTACGATTTCAAGTACATTCATCGGAGCGGATTTGGCGAGCAGG 1792
1775 CTGCTGTTCTACGCTTTTTCATGGAATTTCTTCAATTCAGGATTTGAGAAACAG 1834

RESULT 4

US-09-758-269-1
; Sequence 1, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1749)
US-09-758-269-1

Query Match 41.0%; Score 738.2; DB 10; Length 1752;
Best Local Similarity 68.2%; Pred. No. 9.8e-244;
Matches 1063; Conservative 0; Mismatches 478; Indels 18; Gaps 2;

QY 244 AACACTAAACAGATGAATTTGTTCCAGAGAGCGCGGCGAGCGTTGAGACGCGCGGAG 303
DB 193 AATCTCTCCGGCTAAACATCTTCCAGAAAGCGCGGCGATTGCGATCGACGCGCTGAG 252
QY 304 GGTTCTCTGTCAGCCACGAGAGCTACACCGCTTCTTAAACGCTGATCCTAGTGT 363
DB 253 CGTGCATTAATCTCACACGAGCAAGATTCTCACATTTCCAAACCGCTGATCCAGTGT 312
QY 364 CAGATCGCCGGAATTTGCTCGGTGAATGAACAGCCCGCTCCGCGTAAATCTTCCGCTG 423

Db 313 CAGATTCCGGGAATTATTTCCCGGTACCGGAATCTTCCGTCCGGCGAAACCTCACCGTC 372
Qy 424 GTCCGAAAATTCCCGATTCATCAAGAGGTATGTGGCAACGAGCTAACCCACTT 483
Db 373 GAAGGAACAATCCCTGATGCAATGACGGTGTATATATCCGTAAACGGCGGAATCCGATG 432
Qy 484 CACGAGCCGGTGACAGGTGACACCTTCTTCACGGAGACGGTATGGTTTACGCCGTCAAA 543
Db 433 TTTGAGCCACAGCTGGGCACCAATTTATTCAGCGAGACGGAATGGTTTACGCGATTAA 492
Qy 544 TTCGAAACGGTTTACGTAGTACGTTGCCGGTTTACTAGACTAACCGGTTTGTTCAG 603
Db 493 ATAACCAACGGTTTACGTAGTACGCTACCGGTTTACAAAACCGGAGATTGGTTTTCAG 552
Qy 604 GAACTCAATTTGGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTTCAGCGCCACACC 663
Db 553 GAAAACAGTTTGGGTTCGACCGGTTTCCGGAAGCAATCCGAGCTTTCAGGTCACTCG 612
Qy 664 GGTATTGCCGACTCATCTATTTACGCCAGAGCTGACGCCGTATAGTCGACCCGGCA 723
Db 613 GGAATCGCACGTTTATCTGTTTACGCACGTGGGCTTTTGTGCTCTGATCAACACCAA 672
Qy 724 CACGAAACGGGTAGCTTAAGCCGGTTTGTGCTATTTCAATGCCGTTTATTCGCTATG 783
Db 673 AACGGCTCGAGTAGCAACCGCGTTTGGTTTACTTTAATAACCGGCTTTTACGCTATG 732
Qy 784 TCGGAGGATGATTACCTTTACAAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTT 843
Db 733 TCAGAAGACGATTACCGTACCAATTAATAAATTACTCAACCGCGATCTCCAAACCGTT 792
Qy 844 GGTGGTTTCGATTTTGTGGCAATTAAGATCCCAATGATGCCACCGGAAAGTTCGAC 903
Db 793 GGAAGTTTACGATTTTCGAGCGTCACTTAAATCCGCAATGATAGTCCACCGGAACTCGAC 852
Qy 904 CCGGAATCCGGTGAATCTTCCGTTTAAAGTACGACGTCGTTTCAAGGCTTACCTAAAA 963
Db 853 CCGGTTACGAGGAGCTTCACGCTTAAGCTACGACGTCGTTTAAAGAACCTTACCTGAAA 912
Qy 964 TACTTCCGATTTTCACCGGAGGAACTAAATCACCGGACGTCGAGTTCAGCTTGTATCAG 1023
Db 913 TACTTTCAGATTCTCCCGAGCGGCTTAAATTCGCGGAATGGAGATCCCGCTCGAACT 972
Qy 1024 CCAAGATGATCACGATTTCCGATTTACAGAGAACTTCGTCGTCGATCTGACGACGAA 1083
Db 973 CGAGCATGATTCAGATTTCGCTATACGGAGATTTTGGTGATTCCTGTATCAACAA 1032
Qy 1084 GTCTGTTTCAAGCTCCGAGAGATGATCCGCGTGGGTCTCCGGTGGTTTACGACAGAAC 1143
Db 1033 GTCTGTTTCAAGCTCCGAGAGATGATTTCCGGTAAATCTCCGGTTGTTTTCGACGGAGAA 1092
Qy 1144 AAGGTCCGAAAGATTCCGGATTTTACAGAAATACGCGGAAGATTATCGAAATTAAGTGG 1203
Db 1093 AAGGTTTCCGATTTGGGATTAATGCCAAGNCGCAGAGAGCTTCTCAGATTAATCTGG 1152
Qy 1204 ATTGATGTCAGATTGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGAGCCAGAAACA 1263
Db 1153 GTGAATCTCCGAGACGTTCTGTTTTCATCTCTGGAATGCAATGGGAATCCCGGAGAGC 1212
Qy 1264 GATGAAGTCGTGATAGGTCCTGTATGATCTCCACGAGCTCAATTTTCAAGAGTCT 1323
Db 1213 GAGGAGATTGTTGATCGGATCGTGTATGTCGCGCGGGATTAATCTTCAACAGAGA 1272
Qy 1324 GACGAGAATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGAATCA 1383
Db 1273 GACGAGAGCTTGAGAAGCGTTTGTTCGGAGATCAGATAAACCCTCAGAACACGTAACACC 1332
Qy 1384 ACTCGCGTCCGATCATCTCCAAAGAGATCAAGATCAACCTCGAGAGCGGATGGTC 1443
Db 1333 ACGGTCGTTCTGTTGTTTAAACAGGAT-----GTAAATTTAGAGATTGGTATGTTT 1386
Qy 1444 AACAGAAAATGCTCGGCGGTAAAAACCAAAATTCGTTTACTTGGCTTTTACCGAGCCGTGG 1503
Db 1387 AACCGGAACCGGTTAGGAGAGAAACCCGGTTTCGCGTTTTCGCTTATGCTTATCTCTGG 1446

RESULT 5

US-09-758-269-13

; Sequence 13, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 13

; LENGTH: 1815

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1812)

; US-09-758-269-13

Query Match 33.7%; Score 606.6; DB 10; Length 1815;
Best Local Similarity 63.5%; Pred. No. 3.5e-198;
Matches 1003; Conservative 0; Mismatches 549; Indels 27; Gaps 4;

Qy 246 CACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGGCTTGGACGCGCGGAGG 305
Db 240 CAAGAAGCAGCTCAACTTGTTCAGCGCGCGCGCGCGCGCTCGACGCGTTCCAGGA 299
Qy 306 TTTCTCTGT-----CAGCCACGAGAAGCTACACCCGCTTCTTAAACCGGCTGATCCTAG 359
Db 300 AGGTTTCGTGCCAACGTCCTCGAGCGGCCCCACGCGGCTGCCAGCACGCGCCGACCCGC 359
Qy 360 TGTTCAGATCGCGGAAATTTTCTCGGTGAATGAACACCGCGCTCCGCGGTAATCTTCC 419
Db 360 CGTGCAGATCGCGGCAACTTTCGCGCCCTCGGGAGAGAGCGCGCTGACGAGTCTCC 419
Qy 420 GGTGGTCGGAATACTTCCCGATTCCTCAAGAGAGTGTATGTGCGCAACGAGCTAACCC 479
Db 420 CGTCTCGGCGCATCCGCCCTTTCATCGAGGGGTCTACGCGGCAACGCGCCCAACCC 479
Qy 480 ACTTCAGAGCCGTTGACAGGTCAACCTTCTTCGACGAGAGCGGTATGGTTTCAACCCGT 539
Db 480 CTGCTTCGACCCCGTCCGCGGCGACCACTCTTCGACGCGCGACCGCATGTTGTCACGCGCT 539

QY 540 ---CAAATTCGAACACGGTTTCAGCTAGCTACGCTTGCCTGTTTACTCAGACTAACCGGTT 596
Db 540 GGGGATACGAACGGCGCGGAGTCTAGCTTGCCTGCGCTTCAGGAGACCGCGCGCT 599
QY 597 TGTTCAGAAACGTCAATTTGGTTCGACCGTTTCCCAAAGCCATTCGGTTCAGCTTCAGG 656
Db 600 GCGCAGGAGCGCGGATCGCGCGCGCTTTCCTCCAAAGGCCATTCGGCAGCTGCACGG 659
QY 657 CCACACGGTATTGCCGACTCATCTATTCTACCCAGAGCTGCAGCCGGTATAGTCGA 716
Db 660 GCATCCGGGATCGCGCGCTTCCCTGTTCTACGCGCGCGCGGTTCGCGCTTCGTGA 719
QY 717 CCGGACACACGGAACCGGTTCAGCTTCAGCCGGTTTGGTCTATTTCATTCAGCCCGGTTATT 776
Db 720 CCGCTCGCGCGCACCGCGGTGGCCAAACCGCGCGCTCTACTTCAACCGCGCGCTGCT 779
QY 777 GCGTATGTCGGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAA 836
Db 780 CGCCATGTCGGAGGACGACCTCCCTTACCACTCCGCTGCGGTGGGAGCAGCGGACCTCGA 839
QY 837 AACCGTTCGCTGCTGATTTTGTATGGAACAATTAGAAATCCAAATGATTTGCCACCCGAA 896
Db 840 GACCGTCGCGCTACGACTTCAGCGGCGAGCTCGGCTGGCCATGATCGCGCACCCCAA 899
QY 897 AGTCGACCCGGAATCCGGTGAACCTTCGCTTTAAGCTACGAGCTCGTTTCAAAGCCTTA 956
Db 900 GTGGACCCGCGCACCGGGAGCTCCACGCTCAGCTACGACGTCATCAAGAGGCGCGTA 959
QY 957 CTTAAATATCTTCCGATTTCTACCGGACGGAATAAATCACCGGACGTCGAGATTCACT 1016
Db 960 CTTCAAGTACTTCTACTTCAGGCCCGACGCGACCAAGTCCGACGAGCTGGAGATCCGCT 1019
QY 1017 TGATCAGCCAAAGATGATGATTCGGATTTCGGATTACAGAACTTCGCTGCTACCTGA 1076
Db 1020 GGAGCAGCCACGATGATCCAGACTTCGCCATCACCGGAGAACTTCGTGTTGTGCCGA 1079
QY 1077 CCAGCAAGTCGTTTCAAGCTGCGGAGATGATCCGCGTGGTCTCCGCTGGTTTACGA 1136
Db 1080 CCACAGTGTGTTTCAAGCTCCAGGAGATGCTGCGCGCGGTTCGCGCTGCTGA 1139
QY 1137 CAAGAACAGGTCGCAAGATTCGGGATTTTAGAACAAATACGCGGAGATTCATCAACAT 1196
Db 1140 CAAGGAGAACGCTCGCGGTTTCGGCTGCTCCCAAGCACGCGCGGACGCTCGGAGAT 1199
QY 1197 TAAGTGTATGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAAGCTTCGGAAGAGCC 1256
Db 1200 GCGTGGTGGAGCTGCGCGACTGCTTCTGCTTCCAGCTTGAACGCTTCGGAGACGA 1259
QY 1257 AGAAACAGATGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1316
Db 1260 GCGACCGGCGAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1319
QY 1317 CGAGTCTGACGAGATCTCAAGAGTTCCTGCTGAAATCCGCTGAAATCTCAAAACCGG 1376
Db 1320 CGAGTCTGACGAGCTTCGAGAGCTGCTGACGAGATCCGCTGAGCGCGCGACGG 1379
QY 1377 TGAATCAACTCGCGCTCCGATCATCTCCAAACGAGATCAACAGTCAACCTCGAACGAG 1436
Db 1380 CCGCTCACGCGCGCGCTCTGCGC---GCCGTCGACGAGAGAACTCGAGGTGGG 1436
QY 1437 GATGCTCAACAGAAATGCTCGCGCTGAAACCAATTCGCTTACTTGGCTTTAGCGGA 1496
Db 1437 CATGCTGAACCGCAACTGCTGGCGCGGAGAGCGCGGTACGCTACCTCGCGTGGCGGA 1496
QY 1497 GCCGTGGCTTAAAGTCTCAGGATTCGCTAAAGTTCATCTCACTACTGGAGAAATTAAGA 1556
Db 1497 GCCGTGGCCCAAGAGTGGGCTTCGCCAAGGAGGACCTGTCACGCGGCGAGCTCACCA 1556
QY 1557 ACATCTTTACGCGATACGCTTACGAGGAGGAGCTCTGTTCTCCCGGAGAGGAG- 1615
Db 1557 GTTCGAGTACGCGGCGCGGTTTCGCGCGGAGCCCTGCTTGTTCCTCCATGACCGCGC 1616

QY 1616 -----GAGAGGAAGACGAAGGATATATCTCTGTTTCTGTTTCCAGACGAA 1661
Db 1617 CGCGGCCACCGCGCGCGAGACGAGGTTACGTGCTTCTGTTTCCAGACGAGCG 1676
QY 1662 GACNTGGAATCGAGTTTACAGATAGTTAAACCCGCTAGCTTAGAGTTTGAAGCAACCGT 1721
Db 1677 CGCGGCGACGTCGAGCTACTTGTGTCNAATGCGCGCGACATCCGCTGGAGCGCACGCT 1736
QY 1722 TAAATTCCTCAAGGTTCCGTACGATTTTCAAGTATCAATTCATCGAGCGCGATTT 1781
Db 1737 TCAGTTCGCTCCCGCTGCTTCCAGCGCACCTTTCATCAGCGGCGCAGGAGCT 1796
QY 1782 GCGAAGCAGGTCGTGTA 1800
Db 1797 CGAGGCCAGGCGGCTGA 1815

RESULT 6
US-09-758-269-9
; Sequence 9, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-758-269-9

Query Match 29.3%; Score 528.2; DB 10; Length 1734;
Best Local Similarity 61.0%; Pred. No. 4,7e-171;
Matches 920; Conservative 0; Mismatches 563; Indels 24; Gaps 3;

QY 306 TTTCTTTGTCAGCCACGAGAGCTACACCGCTTCTTAAACCGCTGATCTTAGTGTCA 365
Db 240 TATGCTTATTCCTATGAGCAGATCGCCGCTTCTTAAACCGACCGCGCGGTCA 299
QY 366 GATCGCGGAAATTTCTCGGTGAATGAACAGCCGCTCCGCGTAACTCTCGGTGCT 425
Db 300 ATTATCAGGTAACTTCGCTTAAATGAATGATTCGCGTTTCAAGTGGT 359
QY 426 CGGAAACTTCCGATTCATCAAGAGTGTATGTCGCAACGAGCTAACCCACTCA 485
Db 360 TGGTCAGATTCTTCTGCTTAAAGGAGTTTATCATCCGTAAACCGTAAACCTTATGTT 419
QY 486 CGAGCGGTGACAGTCAACCACTTCTTCAGCGGAGACGCTATGTTTCAACCGTCAATT 545
Db 420 TCCGCGGTAGCGGACATCATTTTATTCGCGTACGGAATGATTCACGCGGTAGTAT 479
QY 546 CGAACACCGT---TCAGTACGTACGCTTCCGCTTACTCAGACTAACCGTTGTTCA 602
Db 480 CGGTTTGTATTAACAGGTTAGTTACGCTGCGGTACACTTAAACAAACCGCTTGTCA 539
QY 603 GGAAGCTCAATTGGGTTCAGCGGTTTCCCGCAAGCCATCGGTGAGCTTCAACGCGCAC 662
Db 540 AGAACCCGCGCTTGACGATCGGTTTCCCTTAAACCAATTCGCGAGCTTCAACGCGCATTC 599

QY 663 CGGTATTCGCCGACTCATCTATTCTACGCCAGAGCTCAGCGGATAGTCGACCCGGC 722
DB |||||
QY 600 CGGTCTAGCTCGACTCGCTCTCTTACCGCTCGAGCTGGGATCGGTCTAGTGGACGGAC 659
DB |||||
QY 723 ACACGGAAACCGGTGATAGCTAAACCGGTTTGGTCTATTTCATGCGCGGTTATTGGGTAT 782
DB |||||
QY 660 AGTGGCATGGCGTAGCTAACCGCGGTGGTCTTTCTTTAACGCGGTTATTAGCAT 719
DB |||||
QY 783 GTCGAGAGATGATTTACCTTACCAATTCAGATCACTCCCAATGGAGATTTAAAAACCGT 842
DB |||||
QY 720 GTCAGAAGATGATCTCTCTTACCAAGTGAAGATCGACGCTCAAGGAGATCTTGAGACGAT 779
DB |||||
QY 843 TGGTTCGGTTCGATTTTGTATGGACAATTAGAAATCCCAATGATTTGCCACCCGAAAGTCGA 902
DB |||||
QY 780 CGGACGGTTCGATTCGATGACCAATTCAGTCTCTTCACTGATAGCGATCTTAGGTGA 839
DB |||||
QY 903 CCGGAATCCGGTGAACCTCTTCGCTTTTAAGCTACGACGTGCTTTCAAGGCTTACCTTAA 962
DB |||||
QY 840 CGCGACCAACGAGATCTCCATACACTGAGCTACACGTTTGAAGAAACCTCATCTCAG 899
DB |||||
QY 963 ATACTTCGATTCCTACCGGAGGAACTAAATCACCGGAGCTCGAGATTCAGCTTGATCA 1022
DB |||||
QY 900 GTATCTTAAATTCACACGTGGGAAAGACACGTGACGTGGAGATCACGCTCCCTGA 959
DB |||||
QY 1023 GCCAAGATGATGACGATTTCCGATTTACAGAGAACTTCGTCGCTGACCTGACCAACA 1082
DB |||||
QY 960 ACCAAGATGATTCATGATTTCCGATTAACCGAATTTTGTCTTATACCGGATCAGCA 1019
DB |||||
QY 1083 AGTCGTTTCAAGCTGCCGAGATGATCCGCGTGGTCTCGGTGGTTTAGCACAAGAA 1142
DB |||||
QY 1020 AATGGTATTCAAATTCGGAATGATTCGGGCGGGTCAACCGTTATCTAGTTAAAGA 1079
DB |||||
QY 1143 CAAGGTGCGAAGATTCGGGATTTAGACAAATACCGGAGATTCATCGAATTAAGTG 1202
DB |||||
QY 1080 AAAATGGCGAGATTTGGAGTTTGTCAAGCAGGATCTGACCGGTGGATATAAATG 1139
DB |||||
QY 1203 GATGATGCTCAGATGCTCTGCTTCCATCTCTGGAACGCTTGGAGAGCCGGAAC 1262
DB |||||
QY 1140 GGTGATGATACCGGATTTCTGTTTCCATCTATGGAATGCTGGGAGAGAGAACCGA 1199
DB |||||
QY 1263 AGATGAAG-----TCGTGTGATAGGCTCTGTATGATCTCCACAGACTCAATTT 1313
DB |||||
QY 1200 AGAGGAGAGCCAGTTATGCTGTATCGGTCAATGATGAGCCCAACCGACACGATCTT 1259
DB |||||
QY 1314 CAACGAGTCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTGGAATCTCAAAC 1373
DB |||||
QY 1260 TAGTGAATCAGGAGAACCAACCGGTTGAAATTAAGTGAGATCCGCTTAAACATCGTAC 1319
DB |||||
QY 1374 CGGTGAATCAACTCGCGCTCGATCATCTCCACGAGATCAACAGTCAACCTCGAAGC 1433
DB |||||
QY 1320 AAAAGATCGAACCGTAAGTTATCGTAACC-----GGAGTGAATTTAGAGC 1367
DB |||||
QY 1434 AGGATGGTCAACAGAAACATGCTCGGCGGTAAACCAATTCGCTTACTTGGCTTTAGC 1493
DB |||||
QY 1368 GGGTCAATAAACCGTAGTTACGTGGCGCGGAAAGCCAGTTGCTTTTACATAGCAATAGC 1427
DB |||||
QY 1494 CGAGCGCTGCCTTAAAGTCTCAGATTCGCTAAAGTTGATCTCACTACTGGAGAGTTAA 1553
DB |||||
QY 1428 CGATCTCTGGCCCAATGCAAGTGGCAATTCGAAGGTAGATATACAAACGCGACCGTTTC 1487
DB |||||
QY 1554 GAAACATCTTTACCGGGAATAACCGTTACGGAGAGAGCCCTCTGTTTCTCCCGGGAAGG 1613
DB |||||
QY 1488 AGAGTTTAAATTCGACCGGACCGGTTCCGTGGAGAACCGTGTCTTACCGGAGGAGA 1547
DB |||||
QY 1614 AGGAGGAGAGACGAGGATACATCTCTGTTTCGTTACGAGGAGAGACATGGAATC 1673
DB |||||
QY 1548 AGGAGAGAGAGACAAAGGTTATGTAATGGGTTTGTGAGAGACGAGAGAGAGACGAGTC 1607
DB |||||
QY 1674 GGAGTTACAGATAGTTAAACCGGTTAGCTTAGAGTTTGAAGCAACGTTTAACTTCGCTC 1733
DB |||||
QY 1608 GGAGTTTGGTGGTGCACCGCAGGATATGAAGCAAGTCGCGCGCGTGGCTTGCCTGA 1667
DB |||||
QY 1734 AAGGTTCCGTACGGATTTACCGGTATCATTCATCGGAGCGGATGATTTGGCGNACAGT 1793
DB |||||

RESULT 8

DB 1668 GAGGTACTTATGTTTCCATGGAACGTTCTGAGCGAGATCAGTTGAAAGACACT 1727
QY |||||
DB 1794 CGTGCA 1800
DB |||||
DB 1728 TTTCTGA 1734
DB |||||

RESULT 7

US-09-770-696-172/c
; Sequence 172, Application US/09770696
; Patent No. US20010044940A1
; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krickler, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Urban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE REFERENCE: 2031US (PAPA-020PRV)

; CURRENT APPLICATION NUMBER: US/09/770,696

; CURRENT FILING DATE: 2001-01-26

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 911

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 172

; LENGTH: 200

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(200)

; OTHER INFORMATION: n = A,T,C or G

US-09-770-696-172

Query Match 11.0%; Score 198; DB 9; Length 200;

Best Local Similarity 99.0%; Pred. No. 2.1e-57;

Matches 198; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 792 TGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTTTGGTTCGGTT 851

DB 200 TGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTTTGGTTCGGTT 141

QY 852 CGATTTTGTAGGACAATTAATCCCAATGATGCCACCGGAGTCCACCGGAGATC 911

DB 140 CGATTTTGTAGGACAATTAATCCCAATGATGCCACCGGAGTCCACCGGAGATC 81

QY 912 CGGTGAATCTTCGCTTTAAGCTACGAGCTCGTTTCAAGCCCTTACCTAAAAATCTTCG 971

DB 80 CGGTGAATCTTCGCTTTAAGCTACGAGCTCGTTTCAAGCCCTTACCTAAAAATCTTCG 21

QY 972 ATTCTACCGGACGGAACCTA 991

DB 20 ATTCTACCGGACGGAACCTA 1

RESULT 8

```
US-09-878-574-2543
; Sequence 2543, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2543
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-023-Q1-B1-G10
US-09-878-574-2543

Query Match      10.6%; Score 190.4; DB 10; Length 393;
Best Local Similarity 68.5%; Pred. No. 1.4e-54;
Matches 263; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 981 GGACGGAATAATACCGGAGCTGCGAGATTTCAGCTTGTATCAGCCAAAGATGATGCACGA 1040
DB 7 GGACGAGAGGAATGCGCGGACATAGAAATTCCTTGAACGGCGGAGATGACACGA 66

QY 1041 TTTTCGGATATACAGAACTTCGTGTCGTACTGACGACGAGTTCGTTTCAAGCTGCC 1100
DB 67 CTTTCGGATACGGAGATTTCGTGTGTATCCGGACGAGAGTGTGTTCAGCTCGG 126

QY 1101 GGAGATGATCCGGTGGTTCCTGGTGTGTTTACGACAGAACAGTTCGCGAGATTGG 1160
DB 127 CGAAATGATCAAGAGGAGTCCGCGTGTATCTACGACGCGGAGAGAAATCCCGTTCGG 186

QY 1161 GATTTTAGCAATAATACCGGAGATTTCATGCAATTAAGTGGATTGATCTCCAGATTG 1220
DB 187 AATAGTCCAAAGTACGCTTCGACGCTCCAGCATGCTGTGGTGGATCCGACGAC 246

QY 1221 CTTTCCTCCATCTCTGAAAGCTTGGGAGAGCCAGAACAGATGAAGTCTGCTGAT 1280
DB 247 GTTCTTCTTCCATCTCTGAAAGCTTGGGAGAGCCAGAACAGATGAGTGTGTAAT 306

QY 1281 AGGTCCTGTATGATCCACGACTCAATTTTCACGAGTCTGACGAGAACTTCAAGAG 1340
DB 307 AGTTCGTGATGACACACCATTCATTAATTAATGACAGAGAGAGTGAAG 366

QY 1341 TGTCTGTCTGAAATCCGCTGAA 1364
DB 367 CGTTTAAACAGAAAGTAAGGCTGAA 390

RESULT 9
US-09-758-269-3
; Sequence 3, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXYANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1785)
US-09-758-269-3

Query Match      9.4%; Score 169.6; DB 10; Length 1788;
Best Local Similarity 50.2%; Pred. No. 6.1e-47;
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;

QY 426 CGGAAAACCTTCGGATTTCATCAAGGAGTGTATGTGGCAACGAGTAAACCCACTTCA 485
DB 393 CGGCACTCTTCCACTGTCTACTTAAAGCGCTTACATCCGTAACGGTCCAAATTCACAG 452

QY 486 CGAGCCGGTGACAGGTTCACACTTCTTCGACGAGACGGTATGGTTTCAGCCGTCAAATT 545
DB 453 TCTCCCTCGTGGTCTTACCATCTCTTCGACGGGACGGTATGCTTCACGCCATAAAAT 512

QY 546 CGAACACGGTTAGCTAGCTAGCTTTCGGTGTCTTACTCAGACTAACCGGTTTGTTCAGGA 605
DB 513 CCACAAACGGTAAAGCCACTCTCTGTAGCAGATACGTCAGAGCTTATAAATCAACGTCGA 572

QY 606 ACGTCAATTGGGTGCGACCGGTTTCCCAAAGCCATCGGTGAGCTTCACGGGCCACACCG- 664
DB 573 GAACAAACCGGAGCTCCGGTTATGCTAAAGTGTTCGCGATTCAACGGTGTAAACGGC 632

QY 665 --GTATTGCCGACTCATGCTATTCTACGCGAGAGTGCAGCCGGTATAGTCGACCCGGC 722
DB 633 GTCAGTAGCTCGTGGAGCTTTAAACGGCAGCTAGGGTTTAAACGGGACAGTAAATCCGGT 692

QY 723 ACAGGAAACGGGTAGCTAAACGCCGGTTTGGTCTATTTCAATGGCCGGTTATTGGCTAT 782
DB 693 TAACGGATTTGGTTTACCTTACCAAGTTCAGATCTATCCCAATGAGAGATTTAAACCCGT 752

QY 783 GTCGAGGATGATTTACCTTACCAAGTTCAGATCTATCCCAATGAGAGATTTAAACCCGT 842
DB 753 AGGTGAATCTGATTTACCTTACCGCTACCGCTGCGATTAAACGGAATCAGGAGATATTGAAAC 812

QY 843 TGGTCCGTTCCGTTTGTAGTGAACAATAGATCCACAATGATTGCCACCCGGAAGTCCA 902
DB 813 CGGACGGTACGATTTGACGGGAAATTAAGCGATGATGACAGCTCTCTCTTAACACCGA 872

QY 903 CCCGGAATCCGGTGAATCTTTCGCTTTAAGCTACGAGCTGGTTTCAAGCCCTTACCTAAA 962
DB 873 TCCAAATACCGGAGAAACTTTCGCTTTCCGGTACGGTCCGGTCCA---CCGTTTTTAAC 929

QY 963 ATACTTCCGATTTCTACCGGACGGAATTAATCACCGGACGTCGAGA---TTCAGCTTGA 1019
DB 930 ATATTCCGGTTGATTTCCCGCGGAAAAAACAAGAGAGCTTCCGATATTCTCGATGAC 989

QY 1020 TCAGCCAAACGATGATGCAGATTTCGCGATTACAGAGAACTTCGTCTGCTACTGACCA 1079
DB 990 GTCTCCGCTGTTTCTCTCCATGATTCGCGATCACGAAACGTCACGCGATTTCGACAGAT 1049

QY 1080 GCAAGTCG-----TTTTCAAGCTCCGGAGATGATCCGGGCTGGTCTCCGGTGT 1130
DB 1050 TCAGCTTGGCATGAGGATGAACATGTTGGATTGGTTTCTCGAAGGTGGTTCTCCGGTTGG 1109

QY 1131 TTACGACAAAGAAACAGGTTCGCAAGATTTCGGGATTTTAGACAAATATACCCGGAAGTTCATC 1190
DB 1110 TACTGATACCGAAAAAATCCAGGCTTGGAGTATTCTTAAGTACCGCGGAGATGATGTC 1169

QY 1191 GAAATTAAGTGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGA 1250
DB 1170 GGAGATGAATGGTTCGAAAGTTCCTGGAATTCATTCACGCTATTATTAATGCTTGGGA 1229

QY 1251 AGAGCCAAACAGATGAAGTCGTCGTGATAGGGTCTCTGTATGACTCCACGACACTCAAT 1310
DB 1230 TGAAGATGTGAAACACAGCTGTTTGTGATGACCGGATATTATGTCGATGTAACATAC 1289
```

```
QY 1311 TTCAACAGTCTGACGAGATCTCAAGAGTGTCTGTCTGAAATCCGCCTGATCTCAA 1370
DB 1290 TTGAGAGGATGGATCTGGTTC---ATGCTTTGGTGGAGAGTGAAGATCGATCTCGT 1346
QY 1371 AACCGGTGAATCAATCCGCCGCTCCGATCATCTCCAAACGAAGATCAACAAGTCAACCTCGA 1430
DB 1347 CACCGGGATGTGACAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391
QY 1431 AGCAGGATGTTCAACAGAAATCTCTCGCGGTAAACCAATTCGTTTACTTGGCTTT 1490
DB 1392 TTTCCTGTGATTAATCCCGGCTTTCTCGGAGATGTAGCAGGTACGTTTACGCGGGAT 1451
QY 1491 AGCCAGCGCTGGCCTAAAGTCTCAGGATTCGCTAAAGTTCATCTCACTACTGAGAGAGT 1550
DB 1452 TGGAGATCCGATGCCGAGATCTCCGCTGTGTGAGCTTGTGTCTAAGAGATCG 1511
QY 1551 TAAGAAACATC-----TTTACGGGATTAACCGTTTACGAGAGAGAGCTCT 1595
DB 1512 GGATGATTTAGCGTGGCGGTGAGATGTACGTTTACGTTTACGCGGAGAACCGTT 1571
QY 1596 GTTCTCCCGGAGAGAGGAGG-----GAGGAAGACGAAGATACATCTCTG 1643
DB 1572 TTTCGTAGTAGGATCTCTGTGTAACTCCGAGCGGAGAGATGATGTTTATGTTGAC 1631
QY 1644 TTTCCTTCAACGAGAGAGATGAAATCGGAGTTACAGATAGTAAACGCGTTAGCTT 1703
DB 1632 GTATGTTACGATGAAGTACTGAGAGATCGAAGTTTCTGTGTAGTGAACGCTAAATCGCC 1691
QY 1704 AGAGTTGAA-----GCAACGGTTAACTTCCGTCAAGGTTCCGTACGATTTACGG 1757
DB 1692 GGAGCTTGAATCTGTCGCGCGCTGAGGTTGCCGGAAGGTTCCGTACGATTTCCATGG 1751
QY 1758 TACATTCATCGAGCGGATGATTTGCGGAAGC 1789
DB 1752 GTTATTTGCAAGGAAGTGAACCTTAATAAGC 1783
```

RESULT 10

```
US-09-938-842A-1444
; Sequence 1444, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1444
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1444
```

Query Match 9.4%; Score 169.6; DB 10; Length 1788;

Best Local Similarity 50.2%; Pred. No. 6.1e-47; Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;

```
QY 426 CGAAACCTTCCGATTCATCAAGAGAGTGTATGCGCAACGAGGTAAACCTTCA 485
DB 393 CGGCACTTTCCTACTTAAACGCGCTTACATCCGTAACGTTCCAAATCCACAT 452
```

```
QY 486 CGAGCGGTGACAGGTACACCACTTCTTGACGAGACGGTATGGTTACGCGGTCAAAT 545
DB 453 TCTCCCTCTGTGTCTTACCACTCTCTTGACGCGGACGGTATGCTTTCAGCCATAAAT 512
QY 546 CGAACACGGTTAGCTAGCTAGCTTACGCTTTCAGCTTACAGATAACCGGTTGTTTCA 605
DB 513 CCACACGGTAAAGCCACTCTCTGTAGCAGATACGTCAGAGCTTATAATACAACGTCGA 572
QY 606 AGTCAATTTGGTTCGACGCGTTTCCCAAGCCATCGGTGAGCTTTCAGGCGCACACCG- 664
DB 573 GAAACAAACGAGCTCCGTTTATGCTTAACTGTTTCCGGATTCACCGGTGTAACGCG 632
QY 665 --GTATTGCCGACTCATCTATTACGCCAGAGTGCAGCGGTATAGTCGACCCGCG 722
DB 633 GTCAGTAGCTCGTGGAGCTTTTAAACGAGCTAGGTTTAAACGCGACAGTATAATCCGCT 692
QY 723 ACACGAAACCGGTAGCTAAACGCGGTTTGTCTATTTCAATGCGCGGTTATTTGCTAT 782
DB 693 TAAACGCAATTTGTTTGTAGCTTAAACAGTCTAGCTTTCTTCACTAACCGCTCTCTT 752
QY 783 GTCGAGGATGATTTACCTTACCAAGTTCCAGATCACTCCCAATGGAGATTTAAACCGT 842
DB 753 AGGTGAATCTGATTTACCTTACCGCTCGGATTAACCGAATCAGGAGATATTGAACGAT 812
QY 843 TGTCGTTTCGATTTTGTAGTGAACATTAAGATCCCAATGATTCGCCACCCGAAAGTCA 902
DB 813 CGACCGGTAGCTTTCGACGCGGAAATTTAGCGATGATGACAGCTCATCTTAAACCGA 872
QY 903 CCGGAAATCCGTTGAACTCTTCTGCTTTAAGCTACGAGCTCGTTTCAAGACCTTACCTAA 962
DB 873 TCCAATAACCGGAGAAACTTTGCTTTCCGTTACGCTCGGTTCCA---CCGTTTTTAAC 929
QY 963 ATACTTCCGATTTCTACCGGACGGAATTAATCAACCGGAGCTCGAGA---TTCAGCTTGA 1019
DB 930 ATATTTCCGTTTGTATTCGCGCGGAAACAAAGAGAGCTTCCGATATTTCTCGATGAC 989
QY 1020 TCAGCCAAACGATGATGACGAGATTTCCGATTTACAGAGAACTTCGCTCGTACCTGACCA 1079
DB 990 GTCTCGTCTTCTCCATGACTTCCGATCAGAAACGTCACGCGATTTTCGACAGAT 1049
QY 1080 GCAAGTCG-----TTTTCAAGCTGCGGAGATGATCCGCGGTGGGTCTCCGCTGGT 1130
DB 1050 TCAGCTTGGCATGAGGATGAACATGTTGGATTGCTTCTCGAAGTGGTCTCCGCTGG 1109
QY 1131 TTACGACAGAAACAAGTCCGAGATTCGCGATTTTACAGAAATACGCGAAGATTCATC 1190
DB 1110 TACTGATACGGAATAACTCCAGGCTTGGAGTATCTTAAGTACCGCGAGATGATC 1169
QY 1191 GAACATTAAGTGAATGATGCTCCAGATTCCTTCTGCTTCCATCTCTGAAACGCTTGGGA 1250
DB 1170 GGAGATGAATGGTTCGAACTTCTGAACTTCAATATCAATTCAGCTATTAAATGCTTGGGA 1229
QY 1251 AGACCCAGAAACAGATGAGTCTGCTGATAGGCTCTGATGACTCCACGACTCAAT 1310
DB 1230 TGAAGATGATGGAACAGCGCTGTTTGTATGACCGAATATTTATGCTGATGAACTAC 1289
QY 1311 TTTCACGAGTCTGACGAGAACTCAAGAGTGTCTGTCTGAAATCCGCTGAACTCAA 1370
DB 1290 TTTAGAGAGGATGATCTGCTGTTTCAATGCTGAGAGAGTGAAGATCGATCTCGT 1346
QY 1371 AACCGTGAATCACTCCGCGTCCGATCATCTCCACGAGATCAACAGTCAACCTCGA 1430
DB 1347 CACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391
QY 1431 AGCAGGATGTTCAACAGAAACATGCTCGGCCCTTAAACCAATTCGCTTACTTGGCTTT 1490
DB 1392 TTTGCTGTGATTAATCCCGGCTTTCTCGGAGATGTAGCAGGTACGTTTACCGCGAT 1451
QY 1491 AGCGAGCGCTTGAAGCTTAAGTCTCAGGATTCGTTAAAGTTGATCTCAGTCTGAGAGT 1550
DB 1452 TGGAGATCCGATCCGAGATCTCCGCTGTGTGAGAGTGTGATGTTTAAAGAGATCG 1511
QY 1551 TAAGAAACATC-----TTTACGCGGATTAACCGTTACGAGAGAGAGCTCT 1595
```

Db 1512 GGATGATGTCAGGTGGCCGGTAGAATGACGGTTCAGGTGTTACGGCGGAGAACCGTT 1571
QY 1596 GTTCTCTCCCGGAGAGGAGGA-----GAGGAAGACGAGGATACATCTCTG 1643
Db 1572 TTTTGGTAGTAGGATCTCTGTTAATCCGAGCGGAGGAGGATGATGTTTATGTTGGTAC 1631
QY 1644 TTTCTGTTACGACGAGGAGACATGGAATCCGAGGTACAGATAGTTAACGCGTTAGCTT 1703
Db 1632 GTATGTTTACAGTAAAGTACGTCGAGATCCGAAGTTCTGTTGATGGACGCTAAATCGCC 1691
QY 1704 AGAGGTTGAA-----GCAACGGTTAAACTTCCGTCAAGGGTTCGGTACGGATTTCAAG 1757
Db 1692 GGAGCTTGAATCTCTCGCCCGGTGAGGTTGCCGGAAGGTTCCGTACGGATTCATGG 1751
QY 1758 TACATTCATCGGAGCGGATGTTGGCGAAGC 1789
Db 1752 GTTATTTTCAAGGAAAGTGACCTTAATAGC 1783

RESULT 11

US-09-758-269-7
; Sequence 7, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1614)
US-09-758-269-7

Query Match 7.8%; Score 139.8; DB 10; Length 1617;
Best Local Similarity 49.7%; Pred. No. 1.2e-36;
Matches 441; Conservative 0; Mismatches 437; Indels 9; Gaps 3;
QY 392 ATGAACAGCCGTCGGCGTAATCTTCGGTGGTCCGAAACTTCCCGATTCCATCAAG 451
Db 170 ATGAACCTCTCCCGTCAAGGATCTCCCGTCCATGATTTCTCCGAAATGTTGATG 229
QY 452 GAGTGTATGTGCGCAACGGAGCTAACCCACTTTCAGCGCCGGTACAGGTCAACCTTCT 511
Db 230 GTGAATTTGTGAGGTTGGTCCAAACCCCAAGTTTGTGCTGCTGGATATCACTGGT 289
QY 512 TCGACGGAGAGCGGTATGTTCAAGCGGTCAAAATTCGAACAGGTTCACTAGTACGCTT 571
Db 290 TTGATGGAGATGGGATGATTCATGGGTACGCATCAAGATGGGAAAGCTACTTATGTTT 349
QY 572 GCCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCATTTGGTTCGACCGGTTTTC 631
Db 350 CTCGATATGTTAAGACATCACTGCTTAAAGCAGGAAGAGTTCTTCGGAGCTGCCAAATCA 409
QY 632 CCAAGCCATCGGTAGGTTTACGGCCACACCGGTATGCCGCTCATGCTATCTTACG 691
Db 410 TGAA---GATTGGTGACCTTAAGGGTTTTCGATTTGCTTAATGGTCAATGTCCAAAGC 466
QY 692 CCAGAGCTGCAGCCGGTATAGTCGACCCGGCACAGGAAACCGGTGTAGCTAAACCGCGTT 751

Db 467 TGAGAACGAGCTCAAAATATTGGACAACACTTATGGAATGGAATGCAATACAGCAC 526
QY 752 TGCTCTATTTCATGCGCGGTATTGGCTATGTCGGAGGATGATTTTACCTTACCAAGTTC 811
Db 527 TCCTATATCACCATTGGAACACTTCTAGCATTTACAGGAGCAGATAGCCGTAGCTCATCA 586
QY 812 AGATCACTCCCAATGGAGATTTAAACACCGTTGGTCCGTTCCGATTTTGTATGGAATTAG 871
Db 587 AAGTTTTGGAAGATGGAGACCTGCAAACTCTTGGTATAATAGATTATGACAAGAGATTGA 646
QY 872 AATCCACATGATGCGCCACCGGAAAGTCAGCCGGAATCCGCGTAACCTTTCGCTTTAA 931
Db 647 CCACACTCTTCTCATGCTTCAACCAAGTTGACCCGGTTACCGGTGAATGTTTACATTCG 706
QY 932 GCTACGACGTCGTTTCAAGCCCTTACCTAAATPACTTCCGATTTCTCACCGGACGGAATA 991
Db 707 GCTATTCCGATACGCCA---CCTTATCTCATACAGAGTTATCTCGAAAGATGGCATTA 763
QY 992 AATCCCGGAGCTCGAGATTCAGCTTGATCAGCAACAGTATGTCAGGATTTTCGCGATTA 1051
Db 764 TGCATGACCCAGTCCCAATTAATATATCAGAGCTATCATGATGATGATTTTGTCTATTA 823
QY 1052 CAGAGAACTTCGTCTGCTGCTACCTGACCCAGCAAGTCGTTTTTCAAGCTGCCGAGATGATCC 1111
Db 824 CTGAGACTTATGCAATCTTTCATGATCTTCTATGCATTCAGGCCAAAGAAATGGTGA 883
QY 1112 GCGTGGGTCTCCGGT---GGTTACGACAGACAGGTTCGCAAGATTCGGGATTTTAG 1168
Db 884 AAGAGAAAGAAATGATATATCTCATTTGATCCCAAAAAGGCTCGTTTTGGTGTCTTC 943
QY 1169 ACAAATACGCGCAAGATTCATCGAACATTAAGTGGATTGATGCTCCAGATTTGCTTCTGCT 1228
Db 944 CACGCTATGCCAAGATGAATTTATGATTAGTTGGTTGAGCTTCCCACTGCTTTATTT 1003
QY 1229 TCCATCTCTGGAACGCTTGGAGAGCCGCAAGACAGATGAAGTCGTC 1275
Db 1004 TCCCAACGCCAATGCTTGGGAAGAGAGGATGAAGTCGTCCTCATC 1050

RESULT 12

US-09-758-269-17
; Sequence 17, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1614)
US-09-758-269-17

Query Match 7.8%; Score 139.8; DB 10; Length 1617;
Best Local Similarity 49.7%; Pred. No. 1.2e-36;
Matches 441; Conservative 0; Mismatches 437; Indels 9; Gaps 3;
QY 392 ATGAACAGCCGTCGGCGTAATCTTCGGTGGTCCGAAAACTTCCCGATTCATCAAG 451


```

; LENGTH: 367
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-791-402

Query Match          3.4%; Score 60.8; DB 9; Length 367;
Best Local Similarity 50.9%; Pred. No. 9.7e-10;
Matches 171; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 520 GACGGTATGTTTACCGCGTCAAAATTCGAACACGGTTTACGTACGTACGCTTCCGGTTT 579
Db 20 GTCCGTATGTTTACCGCGTCAAAATTCGAACACGGTTTACGTACGTACGCTTCCGGTTT 79
QY 580 ACTCAGACTAACCGGTTTGTTCAGGAAGTCAATTCGGTTCGACCGGTTTTCGCCAAAGCC 639
Db 80 GTCAGACTTATAAATACACGTCGAGAACACACCGGAGCCCGGTTATGCTTACGTCG 139
QY 640 ATCGGTGAGTTTACCGGCCACACCG---GTATTCGCGGACTCATGTATTTACGCCAGA 596
Db 140 TTTTCGGATTCAACGGTGTACCGGCTCAGTAGCTCGTGGAGCTTTAAACGGCAGCTAGG 199
QY 697 GCTGCAGCCGTATATCGACCGGCACACGCGGTAGCTAACCGCGGTTTGGTC 756
Db 200 GTTTTACCGGACAGTATAATCCGGTTAACGGGATTTGGTTAGCTAATACGAGTCTAGCT 259
QY 757 TATTTCAATGCGCGGTTATTTGGCTATGTCGGAGGATGATTTACCTTACCAAGTTCAGATC 816
Db 260 CTCTTCAGTAACCGTCTCTTTGCTTTAGGTGAATCTGATTTACCTTACCGCGGTCGATTG 319
QY 817 ACTCCCAATGGAGATTTAAACCGTTGTCGGTTC 852
Db 320 ACCGAATCCGGAGATATTGAACAGATCGGACGGTAC 355

RESULT 15
US-08-976-063C-21/c
; Sequence 21, Application US/08976063C
; Publication No. US20020182697A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A
; TITLE OF INVENTION: ACID AND THEIR USE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTRA
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,063C
; FILING DATE: 21-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 196 49 655.1 (Germany)
; FILING DATE: 29-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELE:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1518 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (4..1518)
; OTHER INFORMATION: /product=
; OTHER INFORMATION: "Lignostilben-Dioxygenase"
; OTHER INFORMATION: /gene= "lsd"
US-08-976-063C-21

Query Match          2.8%; Score 50; DB 8; Length 1518;
Best Local Similarity 47.5%; Pred. No. 1.4e-05;
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 364 CAGATCCCGGAAATTTTGTCTCCGGTGAATGAACACCGCTCCCGGGTAAATCTTCGGTG 423
Db 1491 CAATTAGTAGGAACACTTCTCCCGACCCCGTATAGGCAGACTTGTTCGATCTAGAGTT 1432
QY 424 GTCGAAAACCTTCCCGATTCCATCAAAAGAGTGTATGTGCGCAACCGAGGTAACCCACTT 483
Db 1431 GACGCGGAATCCCAAAATCAATAAATGGAACGTTCTACCGTAATACGCCAGAGCCTCAA 1372
QY 484 CACGAGCCGGTGAACAGTCAACCACTTCTTCGACGAGACGATGTTTCACGCGTCAAA 543
Db 1371 GTTACCCCGACAAAATTTCCACACCTTCATAGATGAGATGGAATGGCCTCTGCTTCCAC 1312
QY 544 TTCGAACACACGGTTCAGCTAGCTAGCTTGGCGGTTTACTCAGACTAACCGGTTTGTTCAG 603
Db 1311 TTCGAAGATGTCATGTCGACTTCATCAGTCGCTGGGTAAACCCGCTCGATTACGCGCC 1252
QY 604 GAAGTCAATTGGGTGACCGGTTTTCGCCAAACCGATCGGTGAGCTTCACGGCCACACC 663
Db 1251 GAACGACTAGCGCGAAAATCGCTATTGGCATGTACAGAAAACCCCTATACCGACGACACC 1192
QY 664 GGTATTGCCGCACT 677
Db 1191 AGTGTAAGGACT 1178

Search completed: November 16, 2003, 20:26:09
Job time : 600 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 15:19:05 ; Search time 117 Seconds
(without alignments)
6790.510 Million cell updates/sec

Title: US-09-758-269-5
Perfect score: 1800
Sequence: 1 atggcttttttcacggcaac.....tggcgaagcaggctgctgtga 1800

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	55.6	3.1	4403765	3	US-09-103-840A-2
C 2	55.6	3.1	4411529	3	US-09-103-840A-1
C 3	53.6	3.0	4403765	3	US-09-103-840A-2
C 4	53.6	3.0	4411529	3	US-09-103-840A-1
C 5	50	2.8	1518	4	US-08-976-063E-21
C 6	50	2.8	32679	4	US-08-976-063E-1
C 7	40.4	2.2	1953	4	US-09-252-991A-2481
C 8	38.8	2.2	7218	1	US-08-232-463-14
C 9	36.2	2.0	738	4	US-09-252-991A-9920
C 10	36.2	2.0	834	4	US-09-252-991A-10156
C 11	35.6	2.0	36063	4	US-08-311-731A-140
C 12	34.8	1.9	4695	4	US-09-309-572-9
C 13	34.8	1.9	4695	6	5225348-3
C 14	34.8	1.9	8450	4	US-09-091-889A-5
C 15	34.2	1.9	1221	1	US-08-434-881-1
C 16	34.2	1.9	1221	3	US-08-977-771-1
C 17	34.2	1.9	1221	3	US-09-361-773-1
C 18	34.2	1.9	1270	4	US-09-026-408-12
C 19	34.2	1.9	1371	4	US-09-026-408-1
C 20	34.2	1.9	1434	4	US-09-016-434-456
C 21	33.6	1.9	456	4	US-09-328-352-3980
C 22	33.4	1.9	396	4	US-09-702-705-21
C 23	33.4	1.9	396	4	US-09-736-457-21
C 24	33.4	1.9	431	4	US-08-702-705-400
C 25	33.4	1.9	451	4	US-09-736-457-400
C 26	33.4	1.9	454	4	US-09-702-705-1123
C 27	33.4	1.9	454	4	US-09-736-457-1123

C 28	33.4	1.9	457	4	US-09-389-681-297	Sequence 297, App
C 29	33.4	1.9	457	4	US-09-620-405B-297	Sequence 297, App
C 30	33.4	1.9	457	4	US-09-339-338-297	Sequence 297, App
C 31	33.4	1.9	457	4	US-09-433-826B-297	Sequence 297, App
C 32	33.4	1.9	457	4	US-09-604-287A-297	Sequence 297, App
C 33	33.4	1.9	788	3	US-08-991-789A-177	Sequence 177, App
C 34	33.4	1.9	788	4	US-09-062-451-177	Sequence 177, App
C 35	33.4	1.9	788	4	US-09-598-326-177	Sequence 177, App
C 36	33.4	1.9	788	4	US-09-289-188-177	Sequence 177, App
C 37	33.4	1.9	1047	2	US-08-494-907-11	Sequence 11, Appl
C 38	33.4	1.9	1047	5	PCT-US96-10986-11	Sequence 11, Appl
C 39	33.4	1.9	1753	6	5225348-2	Patent No. 5225348
C 40	33.4	1.9	3680	2	US-08-494-907-1	Sequence 1, Appl
C 41	33.4	1.9	3680	5	PCT-US96-10986-1	Sequence 1, Appl
C 42	33.4	1.9	5076	2	US-08-494-907-2	Sequence 2, Appl
C 43	33.4	1.9	5076	5	PCT-US96-10986-2	Sequence 2, Appl
C 44	33.4	1.9	6170	2	US-08-494-907-4	Sequence 4, Appl
C 45	33.4	1.9	6170	5	PCT-US96-10986-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match	3.1%	Score 55.6;	DB 3;	Length 4403765;
Best Local Similarity	53.7%	Pred. No. 1.6e-05;		
Matches 115;	Conservative 0;	Mismatches 99;	Indels 0;	Gaps 0;
QY	419	CGGTGTCGGAACACTTCCGATTCATCAAGAGTATGTGCGCAACGGAGCTAAC	478	
Db	1018607	CGGTGACAGCGAAGTCTCCCGCGACCTGACGCGCATCTACCTGCGCACACCGAGAACC	1018548	
QY	479	CACCTTCAGCGCGGTGACAGGTTCACACTTTCGACGAGAGCGGTATGTTTCACGCCG	538	
Db	1018547	CGCTACACCGCGCATTCGCGACCTACCCCTTCGATGCGGACGGCATGATCATGTCG	1018488	
QY	539	TCAAATTCGAACACGGTTCAGCTAGCTTACGCTTCCCGGTTTACTAGACTAACCGGTTG	598	
Db	1018487	TCGCTTTCGGTGATGGAAGAGCTTCTACCGAACCGATTATTTCGACCGATGGATTCT	1018428	
QY	599	TTCAAGAACGTCATTTGGTTCGACCGGTTTTC	632	
Db	1018427	TGGCCGGAACGAGGCGCGGCGCTGTGGCC	1018394	

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A

Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 3.1%; Score 55.6; DB 3; Length 4411529;
Best Local Similarity 53.7%; Pred. No. 1.6e-05;
Matches 115; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 419 CGGTGGTGGAAACTTCCGATTCATCAAGGAGTGTATGTGCGCAACGGAGTAACC 478
DB 1018605 CGGTGACAGGGGAGTCCCGCCGACCTGGACGGCATCTACCTGCGCAACACCGGAGAAC 1018546

QY 479 CACTTACAGAGCGGTGACAGTTCACACATCTTTCGACGGAGCGGTATGTTTCACGCCG 538
DB 1018545 CGCTACACCGGCATTCGCGACCTACACCCCTTCGATGGCAGCGCATATCATGTGCG 1018486

QY 539 TCAATTTGACACCGTTCAGTACGTTTCGCGTTTACGACGTAACCGGTTTG 598
DB 1018485 TCGGTTTCGGTATGGAAGACCTTACCGCAACCGGATTTATTCGACCGATGATTCCT 1018426

QY 599 TTCAGGAACGTCATTTGGTGTGACCGGTTTCCC 632
DB 1018425 TGGCCGAGACGAGCGCGCGCGCTGTGCCC 1018392

RESULT 3
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.0%; Score 53.6; DB 3; Length 4403765;
Best Local Similarity 49.0%; Pred. No. 7.1e-05;
Matches 143; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 820 CCCAATGGAGATTTAAACCGTGGTGGTTCGATTCGATTCGACAAATAGATCCACA 879

Db 752372 CTCACCGATGAGCTGGACACCGTGGACCTGTGACTTCGACGGCACCCCTGCACGGCGGT 752431
QY 880 ATGATTGCCACCGGAAGTCGACCCGGATCCGGTGAACCTCTTCGCTTTAAGCTACGAC 939
Db 752432 TACACCGCCCATCGCAGCGGTGATCCGACACGGGTGAATCGACCGGTGCTTACTCG 752491
QY 940 GTGCTTTCAAAGCCTTACCTAAATATCTCCGATTTCTCACCGGACGGAACTAAATCACCG 999
Db 752492 TTCGCCCGCGGACACAGAGTGCAGTACTCGGTGATCGGACCGGACGACACGCTGCTCG 752551
QY 1000 GACGTGAGATTGAGCTTGTATCAGCAACGATGATCAGGATTCGCGGATTCACAGAAC 1059
Db 752552 ACGTTGATATCGAGGTGGCGGATCGCGATGATCGCAGCTTCTCCCTGACCCACAC 752611
QY 1060 TTGCTGCTGATCTACCTACGACGAAAGTCGTTTCAAGCTCCGCGAGATGATCC 1111
Db 752612 TACGTGATGATCTACGACCTGCGCGGTGACCTTCGACCCAAATCAGGTGCTGC 752663

RESULT 4
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 3.0%; Score 53.6; DB 3; Length 4411529;
Best Local Similarity 49.0%; Pred. No. 7.1e-05;
Matches 143; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 820 CCCAATGGAGATTTAAACCGTGGTGGTTCGATTCGATTCGACAAATAGATCCACA 879
Db 750412 CTCACCGATGAGCTGGACACCGTGGACCTGTGACTTCGACGGCACCCCTGCACGGCGGT 750471
QY 880 ATGATTGCCACCGGAAGTCGACCCGGATCCGGTGAACCTCTTCGCTTTAAGCTACGAC 939
Db 750472 TACACCGCCCATCGCAGCGGTGATCCGACACGGGTGAATCGACCGGTGCTTACTCG 750531
QY 940 GTGCTTTCAAAGCCTTACCTAAATATCTCCGATTTCTCACCGGACGGAACTAAATCACCG 999
Db 750532 TTCGCCCGCGGACACAGAGTGCAGTACTCGGTGATCGGACCGGACGACACGCTGCTCG 750591
QY 1000 GACGTGAGATTGAGCTTGTATCAGCAACGATGATCAGGATTCGCGGATTCACAGAAC 1059
Db 750592 ACGTTGATATCGAGGTGGCGGATCGCGATGATCGCAGCTTCTCCCTGACCCACAC 750651
QY 1060 TTGCTGCTGATCTACCTACGACGAAAGTCGTTTCAAGCTCCGCGAGATGATCC 1111
Db 750652 TACGTGATGATCTACGACCTGCGCGGTGACCTTCGACCCAAATCAGGTGCTGC 750703

RESULT 5
US-08-976-063E-21/c
Sequence 21, Application US/08976063E
Patent No. 6524831
GENERAL INFORMATION:
APPLICANT: Steinduchel, Alexander

APPLICANT: Priefert, Horst
APPLICANT: Rabenhorst, Jürgen
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
FILE REFERENCE: Bayer-9998-CAO
CURRENT APPLICATION NUMBER: US/08/976,063E
CURRENT FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
PRIOR FILING DATE: 1996-11-29
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 1518
TYPE: DNA
ORGANISM: not required under old rule
US-08-976-063E-21

Query Match 2.8%; Score 50; DB 4; Length 1518;
Best Local Similarity 47.5%; Pred. No. 6.4e-06;
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 364 CAGATGCGCGGAAATTTTCTCCGTTGAATGAACAGCGGTCGCGCGTAACTTCCCGTG 423
Db 1491 CAATTAGTAGGAACACTTCTCCCGCCCGTATAGAGCGAGACTTGTTCGATCTAGAGTT 1432
QY 424 GTCGGAATACTTCCGATTCATCAAGAGGTATGTGCGCAACGCGAGCTAACCCACTT 483
Db 1431 GACGGGAAATCCCAAAATCAATAATGGAAGTTCTACCGTAAACGCCAGAGCTCAA 1372
QY 484 CAGAGCGCGGTGACAGGTCCACACTTCTTCGACGAGAGCGGTATGGTTACGCGGTCAA 543
Db 1371 GTTACCCCAAAAATTCACACCTTCATAGATGGAGATGGAATGGCTCTGCTTCCAC 1312
QY 544 TTCGACACGGTTCAGTACGTTCCGCGTTTACTCAGCTAAACCGGTTTGTTCAG 603
Db 1311 TTCGAAGATGGTTCATGCGACTTCATCAGTCCGTTGTTAAACCGTTCGATTCAGGCC 1252
QY 604 GAACGTCAATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTCAAGCCACACC 663
Db 1251 GAACGACTAGCGGAAATCGCTATTGTCATGTACAGAAACCCCTATACCGAGACACC 1192
QY 664 GGTATGCGCGACT 677
Db 1191 AGTGTAAGGACT 1178

RESULT 6
US-08-976-063E-1/c
Sequence 1, Application US/08976063E
Patent No. 6524831
GENERAL INFORMATION:
APPLICANT: Steinbuechel, Alexander
APPLICANT: Priefert, Horst
APPLICANT: Rabenhorst, Jürgen
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
FILE REFERENCE: Bayer-9998-CAO
CURRENT APPLICATION NUMBER: US/08/976,063E
CURRENT FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
PRIOR FILING DATE: 1996-11-29
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 32679
TYPE: DNA
ORGANISM: Pseudomonas sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (3146)..(3997)
OTHER INFORMATION: gene = "ORF1"

US-08-976-063E-1

Query Match 2.8%; Score 50; DB 4; Length 32679;
Best Local Similarity 47.5%; Pred. No. 4.6e-05;
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 364 CAGATGCGCGGAAATTTTCTCCGTTGAATGAACAGCGGTCGCGCGTAACTTCCCGTG 423
Db 16783 CAATTAGTAGGAACACTTCTCCCGCCCGTATAGAGCGAGACTTGTTCGATCTAGAGTT 16724
QY 424 GTCGGAATACTTCCGATTCATCAAGAGGTATGTGCGCAACGCGACTAACCCACTT 483
Db 16723 GACGGGAAATCCCAAAATCAATAATGGAAGCTTCTACCGTAAATACGCCAGAGCTCAA 16664
QY 484 CAGAGCGCGGTGACAGGTCAACACTTCTTCGACGAGAGCGGTATGTTTCAACGCGTCAA 543
Db 16663 GTTACCCCAAAAATTCACACCTTCATAGATGGAGATGGAATGGCTCTGCTTCCAC 16604
QY 544 TTCGAACACGGTTCAGTACGTTTCCCGGTTTACTCAGACTAACCGGTTTGTTCAG 603
Db 16603 TTCGAAGATGGTTCATGCGACTTCATCAGTCCGTTTAAACCCGCTCGATTCACGGCC 16544
QY 604 GAACGTCAATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTCAAGCCACACC 663
Db 16543 GAACGACTAGCGGAAATCGCTATTGTCATGTACAGAAACCCCTATACCGAGACACC 16484
QY 664 GGTATGCGCGACT 677
Db 16483 AGTGTAAGGACT 16470

RESULT 7
US-09-252-991A-2481/c
Sequence 2481, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2481
LENGTH: 1953
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2481

Query Match 2.2%; Score 40.4; DB 4; Length 1953;
Best Local Similarity 44.1%; Pred. No. 0.011;
Matches 216; Conservative 0; Mismatches 271; Indels 3; Gaps 1;
QY 273 AGCGCGCGCGGACGCGGTTGGACGCGGCGGAGGTTTCTTGTTCAGCCAGGAGCTACA 332
Db 640 AGCGAGGATGCGCGACCGCCCGCGCGGAGGTGGCTTGGAGCATCACCGGGTAGCGA 581
QY 333 CCGCTTCTCTAAAACGGGTGATCTTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAA 392
Db 580 TCCGCTCGGCTTCGCGCAGGCGCTCGCGGAGATCGCGGAGCTTGGCTTCGCTCGCGGG 521
QY 393 TGAACAGCGCGTCCGGGTATCTTCGCGTGTTCGGAATACTTCCGATTCATCAAGG 452
Db 520 TGCAGGCGACCGCGCGCGGATCATGTGCGCGGGCTTCGGTCTTGTTCGCCATCGGC 461
QY 453 AGTGTATGTGCGCAACGGAGCTAACCCACTTCA---CGAGCGGTGACAGTCACTT 509
Db 460 GGATCATTGCGCGCAGCGGCGGATGAACCTGATACCGCGCTCGCGCGAGATTTCCGCCA 401

510	CTTCAGACGGAGACGGTATGTTTCAGCGCGTCMAATTCGAACACGGTTTCAGCTAGCTACGC	569
400	GCTCGGCATTTTCGAGAGGAAGCATATAGCCGGGTCGAGGGCGTCGACGCCACTTTTCCA	341
570	TTGCGCGTTTACTCAGACTAACCGGTTTGTTTCAGAGAACGTCAAATTTGGGTGCAACCGGTTTT	629
340	CGCCAGAGTTTCAACAGCGCGCGGGTTGAGGTAGCCGGCCAGCGGTCCGGGCCGATCGC	281
630	CCCCAAAGCCATTCGGTGAAGTTTCACGGCCACACCGGTATTTGCCGAGCTCATGCTATTCTA	689
280	TGTGGGCTTCGTGCGCGGCTTGACGTGCAGCGGTGGCGATCGGCTTCGGAGTAGACGG	221
690	CGCCAGAGCTGCAGCGCGGTATAGTCGACCCGGGCACAGGAACCGGTGTAGCTAACGCCGG	749
220	CCACGAGCGGATGCCATCTCGCGCAGGCGGCACGATGCGGAGCGGCAATCTCGCCGC	161
750	TTTGGTCTAT	759
160	GGTTGGCGAT	151
Db		
Qy		

RESULT 8

US-08-232-463-14
; Sequence 14, Application US/08232463

Patent No. 5670367
 GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEIFELINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 City: Alexandria
 STATE: VA
 COUNTRY: USA
 Zip: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463

Query Match 2.2%; Score 38.8; DB 1; Length 7218; GC 41.1
Best Local Similarity 7.9%; Pred. NO. 0.087;
Matches 16; Conservative 112; Mismatches 74; Indels 0; Gaps 0;

RESULT 9

US-09-252-991A-9920/c
; Sequence 9920, Application US/09252991A

```

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9920
; LENGTH: 738

```

```
Query Match      2.0%; Score 36.2; DB 4; Length 738;
Best Local Similarity 53.1%; Pred. No. 0.14;
Matches 77; Conservative 0; Mismatches 68; Indels
```

393	Qy	TGAACAGCCCGTCCGGCGTAACTCTTCGGGTGTCGGAACACTCCCGATTCCATCAAGG	452
601	Db	TGAACAGCGCGTGTCTGTTGAGCATGTAGTCGAAGCCCGATTGCGCGCGCAGGCCCGAGG	542
453	Qy	AGTGTATGTGGCAACCGAGCTAACCCOACTTCACGAGCCGGTGACAGGTCAACACTTCTT	512
541	Db	AGTCCTGCAGGCTTCATGCTCTGAAACCCCTGGCCCTTGGGTTGTGTGGGAGGCTTCTGT	482
513	Qy	CGACGGAGACGGTATGGTTCAACGC	537
481	Db	CGAGAAAGGTGCTGTTAGTTCAACGCC	457

RESIST. 70

RESULTS TO
US-09-252-991A-10156
; Sequence 10156, Application US/09252991A

Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 10156
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10156

Query Match
Best Local Similarity 2.0%; Score 36.2; DB 4; Length 834;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 393 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGTCGGAATACTTCCCGGATTCATCAAAAG 452
Db 249 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGTCGGAATACTTCCCGGATTCATCAAAAG 308

QY 453 ACTGTATGTGCGCAACGAGCTAACCCACTTACAGAGCGGTGACAGGTACCACTTCTT 512
Db 309 AGTCCTGCACTTCACTGCTGCTGAACCCCTGCGCTTCCGGTGTGTCGGAATACTTCCCGG 368

QY 513 CCACGAGAGCGGTATGTTTCAGGCC 537
Db 369 CGAAGAGGTGCTGTAGTTCAGGCC 393

RESULT 11
US-08-311-731A-140
; Sequence 140, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae
US-08-311-731A-140

Query Match
Best Local Similarity 2.0%; Score 35.6; DB 4; Length 36063;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

; SEQ ID NO 10156
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10156

Query Match
Best Local Similarity 2.0%; Score 36.2; DB 4; Length 834;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 393 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGTCGGAATACTTCCCGGATTCATCAAAAG 452
Db 249 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGTCGGAATACTTCCCGGATTCATCAAAAG 308

QY 453 ACTGTATGTGCGCAACGAGCTAACCCACTTACAGAGCGGTGACAGGTACCACTTCTT 512
Db 309 AGTCCTGCACTTCACTGCTGCTGAACCCCTGCGCTTCCGGTGTGTCGGAATACTTCCCGG 368

QY 513 CCACGAGAGCGGTATGTTTCAGGCC 537
Db 369 CGAAGAGGTGCTGTAGTTCAGGCC 393

RESULT 12
US-09-309-572-9
; Sequence 9, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 4695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-309-572-9

Query Match
Best Local Similarity 1.9%; Score 34.8; DB 4; Length 4695;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 340 CCTAAACAGCGCTGATCTAGTGTTCAGATCCCGGAAATTTCTCCGGTGAATGAACAG 399
Db 3386 CTTGATTGGATTGCCACACGCGCTCATTCGATGCAAGTTTCTGAGCTGAAGAAAG 3445

QY 400 CCGTCCGCGGTAATCTTCCGGTGTGCGGAAACCTTCCGATTCATCAAGCAGTGAT 459
Db 3446 ATTGATCGCGTCTCTGTTAAAGCTGGAAGATGGCCCTAAATCTTGAATCTGTGAT 3505

QY 460 GTGCGCAACGAGCTAACCCACTTACAGAGCGGTGACAGGTCCACCTTTTCGACGGA 519
Db 3506 GCTGCCATTGTTGATATGTTCTCTGCAAGCCCATGTGTGAGAGCTTCTCAGACTAT 3565

QY 520 GACGGTATGTTTACGCCGCTCAATT 545
Db 3566 CCACCTTTGGTAAGGATGACTACTT 3591

RESULT 13
US25348-3
; Patent No. 525348
; APPLICANT: HAGATA, SHIGEKAZU; SUGANO, SUMIO; KIM, DONG W.;
; UETSUKI, TAICHI; KAZIRO, YOSHITO
; TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
; CONTAINING THE DNA FRAGMENT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,823
; FILING DATE: 08-DEC-1989
; SEQ ID NO: 3
; LENGTH: 4695
US25348-3

Query Match
Best Local Similarity 1.9%; Score 34.8; DB 6; Length 4695;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 13:15:00 ; Search time 6756 Seconds
(without alignments)
10899.545 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atggctcttttcacggaac.....tggcgaagcaggtcgtgtga 1800

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.to.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rat.*
36: em.htg.rod.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1800	100.0	1800	6	AX148306	AX148306 Sequence
2	1800	100.0	1800	6	BD017431	BD017431 Transgeni
3	1800	100.0	3869	8	AB026549	AB026549 Arabidops
4	1800	100.0	52332	8	AB028617	AB028617 Arabidops
5	1798.4	99.9	2331	8	AY056255	AY056255 Arabidops
6	862.8	47.9	95769	6	AX652128	AX652128 Sequence
7	862.8	47.9	95769	6	AC013430	AC013430 Genomic s
8	824.6	45.8	1818	6	AX148316	AX148316 Sequence
9	824.6	45.8	1818	6	BD017436	BD017436 Transgeni
10	824.6	45.8	2171	8	LENEOXANT	297215 Lycopersico
11	824.6	45.8	19018	8	LESA439079	AJ439079 Lycopersi
12	809.4	45.0	2164	8	STU276244	AJ276244 Solanum t
13	798.4	44.4	1839	6	AX148312	AX148312 Sequence
14	798.4	44.4	1839	6	BD017434	BD017434 Transgeni
15	798.4	44.4	2432	8	AB030293	AB030293 Vigna ung
16	791.6	44.0	2398	8	AF190462	AF190462 Phaseolus
17	750.8	41.7	72058	8	AC074176	AC074176 Arabidops
18	745	41.4	2357	8	AB080192	AB080192 Pisum sat
19	738.2	41.0	1752	6	AX148302	AX148302 Sequence
20	738.2	41.0	1752	6	BD017429	BD017429 Transgeni
21	738.2	41.0	110102	8	ATF28J12	AL021710 Arabidops
22	738.2	41.0	194143	8	ATCHRIV48	AL161548 Arabidops
23	696.8	38.7	2153	8	AF224672	AF224672 Persea am
24	693.8	38.5	2034	8	AB080193	AB080193 Pisum sat
25	665	36.9	2310	8	AF224671	AF224671 Persea am
26	655.8	36.4	175158	2	AC120531	AC120531 Oryza sat
27	655.8	36.4	181159	2	CNS071P7	AL713390 Oryza sat
28	645	35.8	131689	8	CNS08CB9	AL845347 Oryza sat
29	645	35.8	133983	8	CNS08CBP	AL731885 Oryza sat
30	606.6	33.7	1815	6	AX148314	AX148314 Sequence
31	606.6	33.7	1815	6	BD017435	BD017435 Transgeni
32	606.6	33.7	2498	8	ZMU95953	U95953 Zea mays vi
33	573.8	31.9	149097	2	AP005632	AP005632 Oryza sat
34	528.2	29.3	1734	6	AX148310	AX148310 Sequence
35	528.2	29.3	1734	6	BD017433	BD017433 Transgeni
36	528.2	29.3	78921	8	AB028621	AB028621 Arabidops
37	498.4	27.7	82697	8	ATT9A21	AL021713 Arabidops
38	314.6	17.5	602	8	AF399655	AF399655 Lactuca s
39	201	11.2	365	11	BX295824	BX295824 Arabidops
40	195.2	10.8	2016	8	AB080194	AB080194 Pisum sat
41	186.8	10.4	80866	2	AC144759	AC144759 Medicago
42	186.8	10.4	120055	8	AC123596	AC123596 Medicago
43	177.4	9.9	1971	8	MD293765	Z93765 M.domestica
44	176	9.8	120055	8	AC123596	AC123596 Medicago
45	171.2	9.5	1788	8	BT008785	BT008785 Arabidops

ALIGNMENTS

RESULT 1	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
LOCUS	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
DEFINITION	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
ACCESSION	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
VERSION	AX148306.1	GI:14347193	1800 bp	DNA	linear	PAT 08-JUN-2001
KEYWORDS	Arabidopsis thaliana (thale cress)					
SOURCE	Arabidopsis thaliana					
ORGANISM	Arabidopsis thaliana					
REFERENCE	1	Arabidopsis thaliana (thale cress)				
AUTHORS	luchi,S., Kobayashi,M. and Shinozaki,K.					
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene					

Qy	661	ACCGGTATTGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCCGCTATAGTCGACCCG	720
Db	661	ACCGGTATTGCCCGACTCATGCTATTCTACGCCAGAGCTGCAGCCGCTATAGTCGACCCG	720
Qy	721	GCACACGGAAACCGGTGTAGCTAAACGCCGGTTTGGTCTATTTCAATGSCCGGTATTGGCT	780
Db	721	GCACACGGAAACCGGTGTAGCTAAACGCCGGTTTGGTCTATTTCAATGSCCGGTATTGGCT	780
Qy	781	ATGTCGGAGGATGATTTACCTTTACCAAGTTCAGATCACTCCCAATGAGAGATTAAAAACC	840
Db	781	ATGTCGGAGGATGATTTACCTTTACCAAGTTCAGATCACTCCCAATGAGATTTAAAAACC	840
Qy	841	GTTGGTCGGTTCGATTTTGATGAGCAATATAGAAATCCAAATGATGCGCCACCGAAAGTC	900
Db	841	GTTGGTCGGTTCGATTTTGATGAGCAATATAGAAATCCAAATGATGCGCCACCGAAAGTC	900
Qy	901	GACCCGGAATCCCGGTGAACCTCTTCGGTTTAAAGCTACGACGTCGTTTCAAAAGCCTTACCTA	960
Db	901	GACCCGGAATCCCGGTGAACCTCTTCGGTTTAAAGCTACGACGTCGTTTCAAAAGCCTTACCTA	960
Qy	961	AAATACHTCCGATTTCTACCCGAGCGGAACCTAAATTCACCGGACGTCGAGATTACGTTGAT	1020
Db	961	AAATACHTCCGATTTCTACCCGAGCGGAACCTAAATTCACCGGACGTCGAGATTACGTTGAT	1020
Qy	1021	CAGCCACGATGATGCACGATTTTCGGGATTACAGAGAACTTCGTCGTCGTACCTGCACG	1080
Db	1021	CAGCCACGATGATGCACGATTTTCGGGATTACAGAGAACTTCGTCGTCGTACCTGCACG	1080
Qy	1081	CAAGTCGTTTCAAGCTGCCGAGATGATCCGCGGTGGTCTCCGGTGGTTTACGACAAG	1140
Db	1081	CAAGTCGTTTCAAGCTGCCGAGATGATCCGCGGTGGTCTCCGGTGGTTTACGACAAG	1140
Qy	1141	AACAAGTCCCAAGATTCCGGGATTTTACACAAATACCCGGAAGATTTCATCGAAATTAAG	1200
Db	1141	AACAAGTCCCAAGATTCCGGGATTTTACACAAATACCCGGAAGATTTCATCGAAATTAAG	1200
Qy	1201	TGATTTGATGCTCCAGATTGCTTCTGCTTCGATCTCTGGAAACGTTCCGGAAGCCAGAA	1260
Db	1201	TGATTTGATGCTCCAGATTGCTTCTGCTTCGATCTCTGGAAACGTTCCGGAAGCCAGAA	1260
Qy	1261	ACAGATCAAGTCGTCTGATAGAGTCTGTATGACTCCACAGACTCAATTTTCAACGAG	1320
Db	1261	ACAGATCAAGTCGTCTGATAGAGTCTGTATGACTCCACAGACTCAATTTTCAACGAG	1320
Qy	1321	TCTGACAGAAATCTCAAGATGTCTGTCTGAAATCCGCTGTAATCTCAAAACCGGTGAA	1380
Db	1321	TCTGACAGAAATCTCAAGATGTCTGTCTGAAATCCGCTGTAATCTCAAAACCGGTGAA	1380
Qy	1381	TCAACTCCCGTCCGATCATCTCCAAACGAAAGTCAACAAAGTCAACCTCGAAGCAGGATG	1440
Db	1381	TCAACTCCCGTCCGATCATCTCCAAACGAAAGTCAACAAAGTCAACCTCGAAGCAGGATG	1440
Qy	1441	GTCAACAGAAACATGCTCCGCGGTAAACCAAAATTCGTTACTTTGGCTTTAGCCGAGCG	1500
Db	1441	GTCAACAGAAACATGCTCCGCGGTAAACCAAAATTCGTTACTTTGGCTTTAGCCGAGCG	1500
Qy	1501	TGGCTTAAAGTCTCAGGATTTCGTTAAAGTTGATCTCACTACTGGAGAAGTTTAAAGAACAT	1560
Db	1501	TGGCTTAAAGTCTCAGGATTTCGTTAAAGTTGATCTCACTACTGGAGAAGTTTAAAGAACAT	1560
Qy	1561	CTTTTACCGGATACCGTTACGGAGGAGACCTCTGTTTCTCCCGGAGAGGAGGAGAG	1620
Db	1561	CTTTTACCGGATACCGTTACGGAGGAGACCTCTGTTTCTCCCGGAGAGAGGAGGAGAG	1620
Qy	1621	GAAGACGAAGGATATACCTCTGTTTCGTTTACGACGAGAGACATGGAAATCGGAGTTA	1680
Db	1621	GAAGACGAAGGATATACCTCTGTTTCGTTTACGACGAGAGACATGGAAATCGGAGTTA	1680
Qy	1681	CAGATAGTTAACCGGTTAGCTTAGAGGTTGAGCAACGGTTTAACTTCCTCGTCGAGGGTT	1740
Db	1681	CAGATAGTTAACCGGTTAGCTTAGAGGTTGAGCAACGGTTTAACTTCCTCGTCGAGGGTT	1740

```

QY 1741 CGGTACGGATTTACGGTACATTCATCGAGCGCGATGATTGGGAAACAGGTGCTGTGA 1800
Db |||||
QY 1741 CGGTACGGATTTACGGTACATTCATCGAGCGCGATGATTGGGAAACAGGTGCTGTGA 1800
Db |||||

RESULT 2
LOCUS BD017431 1800 bp DNA linear PAT 27-AUG-2002
DEFINITION Transgenic plant using neozanthine cleaving enzyme gene.
ACCESSION BD017431
VERSION BD017431.1 GI:22558607
KEYWORDS JP 2001258579-A/3.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1800)
REFERENCE Luchi, K., Kobayashi, M. and Shinozaki, K.
AUTHORS Transgenic plant using neozanthine cleaving enzyme gene
TITLE Patent: JP 2001258579-A 3 25-SEP-2001;
JOURNAL THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Arabidopsis thaliana (thale cress)
PN JP 2001258579-A/3
PD 25-SEP-2001
PF 11-JAN-2001 JP 2001003476
PI KIYOSHI TUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC
C12N15/09, A01H5/00, C12N5/10, C12N9/02, C12N5/00, C12N5/00 CC
Transgenic plant using neozanthine cleaving enzyme gene FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..1800
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 458 a 464 c 439 g 439 t
ORIGIN
Query Match 100.0%; Score 1800; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCTTTCACGGCAACGCTGCGGTTCTCGGAGATGCTTGGTGGCAATCATACT 60
Db |||||
QY 1 ATGGCTCTTTCACGGCAACGCTGCGGTTCTCGGAGATGCTTGGTGGCAATCATACT 60
Db |||||
QY 61 CAGCGCCATTATCGCTTCTCAAAGCTCCGACTTGGATTTAGTCTCCTTACCTATG 120
Db |||||
QY 61 CAGCGCCATTATCGCTTCTCAAAGCTCCGACTTGGATTTAGTCTCCTTACCTATG 120
Db |||||
QY 121 GCCAGTCGTGCACAGTAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGCTTT 180
Db |||||
QY 121 GCCAGTCGTGCACAGTAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGCTTT 180
Db |||||
QY 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCATTTGTTTAAAGCCCAAGCAAGAA 240
Db |||||
QY 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCATTTGTTTAAAGCCCAAGCAAGAA 240
Db |||||
QY 241 TCCAAACACTAAACAGATGAATTTGTTCCAGAGAGCGGCGCGAGCGTGGACGCGCG 300
Db |||||
QY 241 TCCAAACACTAAACAGATGAATTTGTTCCAGAGAGCGGCGCGAGCGTGGACGCGCG 300
Db |||||
QY 301 GAGGGTTTCCTTGTGACGACACAGAGAGCTACACCGGCTTCCTTAAACCGGCTGATCCTAGT 360
Db |||||
QY 301 GAGGGTTTCCTTGTGACGACACAGAGAGCTACACCGGCTTCCTTAAACCGGCTGATCCTAGT 360
Db |||||
QY 361 GTTCAGATCCCGGAATTTTGTCTCGGTGAATGAACACCGCGTCCGGGTAACTCTCCG 420
Db |||||
QY 361 GTTCAGATCCCGGAATTTTGTCTCGGTGAATGAACACCGCGTCCGGGTAACTCTCCG 420
Db |||||
QY 421 GTGGTCGGAACACTTCCCGATTCATCAAGAGGTGATGTGCGCAACGAGCTTACCA 480
Db |||||

```

```

Db 421 GTGGTCGGAACACTTCCCGATTCATCAAGAGGTGATGTGCGCAACGAGCTAACCCA 480
QY 481 CTTTCAGAGCGGTGACAGGTACACACTTCTTTCGACGGAGAGCGTATGTTTACGCGCTC 540
Db 481 CTTTCAGAGCGGTGACAGGTACACACTTCTTTCGACGGAGAGCGTATGTTTACGCGCTC 540
QY 541 AAATTTCGAACACCGGTTACAGTACGTTCGCGGTTTACTCAGACTAACCGGTTTGT 600
Db 541 AAATTTCGAACACCGGTTACAGTACGTTCGCGGTTTACTCAGACTAACCGGTTTGT 600
QY 601 CAGGAAACGTCATATGGGTGACCGGTTTCCCAAGGCCATCGGTAGCTTACCGGCCAC 660
Db 601 CAGGAAACGTCATATGGGTGACCGGTTTCCCAAGGCCATCGGTAGCTTACCGGCCAC 660
QY 661 ACCGGTATTGCGCGACTCATGTATTCTACGCCAGAGCTGCAGCGGTATAGTCGACCG 720
Db 661 ACCGGTATTGCGCGACTCATGTATTCTACGCCAGAGCTGCAGCGGTATAGTCGACCG 720
QY 721 GCACACGGAACCGGTGTAGCTAACGCGGTTTGGTCTATTTCATAGCGCGTTATTGGCT 780
Db 721 GCACACGGAACCGGTGTAGCTAACGCGGTTTGGTCTATTTCATAGCGCGTTATTGGCT 780
QY 781 ATGTGCGGAGATGATTTTACCTTACCAGTTTCAGATCACTCCCAATGAGATTTAAAAAC 840
Db 781 ATGTGCGGAGATGATTTTACCTTACCAGTTTCAGATCACTCCCAATGAGATTTAAAAAC 840
QY 841 GTTGGTGGTTCGATTTTATGATGACAAATTAGAATCCAAATGATTCGCCACCGGAAAGTC 900
Db 841 GTTGGTGGTTCGATTTTATGATGACAAATTAGAATCCAAATGATTCGCCACCGGAAAGTC 900
QY 901 GACCCGGATCCCGTGAACCTTCGCTTTAAGTACGAGCTGCTTTCAGAGCTTACCTTACCTA 960
Db 901 GACCCGGATCCCGTGAACCTTCGCTTTAAGTACGAGCTGCTTTCAGAGCTTACCTTACCTA 960
QY 961 AAATCTTCCGATTTCTCACGGACGGAACTAAATCAACCGAGCTCGAGATTTCAAGTTGAT 1020
Db 961 AAATCTTCCGATTTCTCACGGACGGAACTAAATCAACCGAGCTCGAGATTTCAAGTTGAT 1020
QY 1021 CAGCAACAGATGATGACGATTTCCGATTCAGAGAACTTCGCTCGTCTACCTGACCCAG 1080
Db 1021 CAGCAACAGATGATGACGATTTCCGATTCAGAGAACTTCGCTCGTCTACCTGACCCAG 1080
QY 1081 CAAGTCGTTTTCAAGCTGCGGAGATGATCCCGGTTGGTCTCCGTTGGTTTACGCAAG 1140
Db 1081 CAAGTCGTTTTCAAGCTGCGGAGATGATCCCGGTTGGTCTCCGTTGGTTTACGCAAG 1140
QY 1141 AACAGTTCGCAAGATTCGGGATTTAGACAAATACGCGAGATTCATCGAACATTAAG 1200
Db 1141 AACAGTTCGCAAGATTCGGGATTTAGACAAATACGCGAGATTCATCGAACATTAAG 1200
QY 1201 TGGATTGATGCTCCAGATTCGTTCTGCTTCCATCTCTGGAACGCTTGGGAAGCCAGAA 1260
Db 1201 TGGATTGATGCTCCAGATTCGTTCTGCTTCCATCTCTGGAACGCTTGGGAAGCCAGAA 1260
QY 1261 ACAGATGAATCGTGTGATGAGGTCTCTGATGATCCACAGATCAATTTTCAACGAG 1320
Db 1261 ACAGATGAATCGTGTGATGAGGTCTCTGATGATCCACAGATCAATTTTCAACGAG 1320
QY 1321 TCTGACGAGATCTCAAGAGTCTCTGTAATCGCTGAATCTCAAAACCGGTGAA 1380
Db 1321 TCTGACGAGATCTCAAGAGTCTCTGTAATCGCTGAATCTCAAAACCGGTGAA 1380
QY 1381 TCAACTCGCGCTCGATTCATCTCCAAACGAGATCAACAAAGTCAACTCGAAGCAGGATG 1440
Db 1381 TCAACTCGCGCTCGATTCATCTCCAAACGAGATCAACAAAGTCAACTCGAAGCAGGATG 1440
QY 1441 GTCAACAGAAACATGTCGCGCGTAAACCAAAATTCGTTTACTTTCGCGGAGCGCG 1500
Db 1441 GTCAACAGAAACATGTCGCGCGTAAACCAAAATTCGTTTACTTTCGCGGAGCGCG 1500
QY 1501 TGGCCTTAAAGTCTCAGGATTCGTTTAAAGTTGATCTCACTACTGAGAGTAAAGAACAT 1560
Db 1501 TGGCCTTAAAGTCTCAGGATTCGTTTAAAGTTGATCTCACTACTGAGAGTAAAGAACAT 1560

```

QY 1561 CTTTACGGCATACCGTTACGGAGAGAGCCTCTGTTTCTCCCGGAGAGGAGAG 1620
 Db 1561 CTTTACGGCATACCGTTACGGAGAGAGCCTCTGTTTCTCCCGGAGAGGAGAG 1620
 QY 1621 GAAGACGAGGATACATCCTCTGTTTCTGTTACGACGAGAGACATGAAATCGAGTTA 1680
 Db 1621 GAAGACGAGGATACATCCTCTGTTTCTGTTACGACGAGAGACATGAAATCGAGTTA 1680
 QY 1681 CAGATAGTTAACCGCGTTAGCTTAGAGTTGAGGTTGAGCAACGGTTAACTTCCGTTAAGGGTT 1740
 Db 1681 CAGATAGTTAACCGCGTTAGCTTAGAGTTGAGGTTGAGCAACGGTTAACTTCCGTTAAGGGTT 1740
 QY 1741 CCGTACGGATTTACGGTACATTCATCGGAGCCGATGATTGCGCAAGCAGTCTGTGTA 1800
 Db 1741 CCGTACGGATTTACGGTACATTCATCGGAGCCGATGATTGCGCAAGCAGTCTGTGTA 1800

RESULT 3

AB026549 3869 bp DNA linear PLN 24-OCT-2001
 LOCUS Arabidopsis thaliana gene for neoxanthin cleavage enzyme, complete cds.

AB026549
 VERSION AB026549.1 GI:16416373
 KEYWORDS neoxanthin cleavage enzyme.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE

1 Iuchi,S., Kobayashi,M. and Shinozaki,K.
 Characterization of neoxanthin cleavage enzyme from Arabidopsis thaliana
 Unpublished
 2 (bases 1 to 3869)
 Iuchi,S. and Shinozaki,K.
 Direct Submission
 TITLE Submitted (22-APR-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.;
 3-1-1 Kouyadai, Tsukuba 305-0074, Japan
 (E-mail:iuchi@rtc.riken.go.jp, Tel:81-298-36-4359)

FEATURES

source

1..3869

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Col."

/db_xref="taxon:3702"

1831..3630

/codon_start=1

/product="neoxanthin cleavage enzyme"

/protein_id="BAB70509.1"

/db_xref="GI:16416374"

/translations="MASFTATAAVSGRWLGNHTQPLSSSSDLSYCSLPMSRV
 TRKLVNSALHTPPLHPKSSNSPAIVVKPKESNTKQNLFORAAALDAEG
 FLVSHKLPPLTADPSVQIAGNFAPVNEQVRNLPVVKGLPDSIKGVYRNANP
 LHPVTVGHFDFDGMVHAKFHEGSAVACRTQNRVQERQLRPFVPAIGELH
 GHTGIARLMFYAAGAGIVDPAHGTGVANAGLVYFNGLLMSDEDLPEVOITPNG
 DLTKVDFDFDGOLESTMAIAHPKVDPESCLEPALSVDVVKPLKYFRSPDGTSPD
 VEIQDQPTWHDHFAITENFVVDQVFKLFEMIRGSGSPVYDKNKVARGILDKY
 ADESSNIRKWDADPCFDFHUNAWBEPEDEVVIGSMTPPDSIFNESDNLKSVLS
 EIRLNKLGSTSRPISNEDQVNLGAEVNMNMLGRKTKFAYLALAEFPWPKVSGFA
 KVDLTGGEVKHLYGNRYGGEPLFLPGEGBEDEGYILCFVHDXTKRSQELQVNAV
 SLEVEATVKLPSPVPYGFHGTFIGADDLAKQVV"

BASE COUNT 1155 a 845 c 773 g 1096 t

ORIGIN

Query Match 100.0%; Score 1800; DB 8; Length 3869;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTTCTTTACGGCAACGGTCGGTTTCGGAGAGTGGTTCGGTCAATCATCT 60

Db 1831 ATGGCTTCTTTACGGCAACGGTCGGTTTCGGAGAGTGGTTCGGCAATCATACT 1890
 QY 61 CAGCCGCCATATPCGTCCTTCTCAAAGCTCCGACTTGAGTTATGTAGTCTCTTACCTATG 120
 Db 1891 CAGCCGCCATATPCGTCCTTCTCAAAGCTCCGACTTGAGTTATGTAGTCTCTTACCTATG 1950
 QY 121 GCCAGTCGTGTACACGTAAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGTCTTT 180
 Db 1951 GCCAGTCGTGTACACGTAAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGTCTTT 2010
 QY 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCAATTTGTTTAAGCCCAAGCAAGAA 240
 Db 2011 CATTTCCCTAAGCAATCATCAAACTCTCCGCAATTTGTTTAAGCCCAAGCAAGAA 2070
 QY 241 TCCAAACACTAAACAGATGAATTTGTTTCCAGAGAGCGCGCGGACGCTTGACACGCGCG 300
 Db 2071 TCCAAACACTAAACAGATGAATTTGTTTCCAGAGAGCGCGCGGACGCTTGACACGCGCG 2130
 QY 301 GAGGTTTCTTGTGACGCCACGAGAGCTACACCCCGCTTCTTAAACCGGCTGATCTAGT 360
 Db 2131 GAGGTTTCTTGTGACGCCACGAGAGCTACACCCCGCTTCTTAAACCGGCTGATCTAGT 2190
 QY 361 GTTCAGATCGCCGGAATTTTGTCTCCGTTGAATGAACGCCGCTCGGCGTATCTTCG 420
 Db 2191 GTTCAGATCGCCGGAATTTTGTCTCCGTTGAATGAACGCCGCTCGGCGTATCTTCG 2250
 QY 421 GTGTCGGAAACCTTCCCGATTTCCATCAAAGGAGTGTATGTGCGCAACGGAGTAAACCA 480
 Db 2251 GTGTCGGAAACCTTCCCGATTTCCATCAAAGGAGTGTATGTGCGCAACGGAGTAAACCA 2310
 QY 481 CTTTCAGAGCGCGTGACAGGTACCACTTCTTCGAGGAGACGAGTATGTTTCACGCGGTC 540
 Db 2311 CTTTCAGAGCGCGTGACAGGTACCACTTCTTCGAGGAGACGAGTATGTTTCACGCGGTC 2370
 QY 541 AAATTCGAACACCGTTTCAGCTAGCTAGCTTGCCTGCTTACTCAGACTAACCCGTTTGT 600
 Db 2371 AAATTCGAACACCGTTTCAGCTAGCTAGCTTGCCTGCTTACTCAGACTAACCCGTTTGT 2430
 QY 601 CAGGAACGCTCAATGGGTGACCGGTTTCCCAAGGCCATCGTGAGCTTACGGCCAC 660
 Db 2431 CAGGAACGCTCAATGGGTGACCGGTTTCCCAAGGCCATCGTGAGCTTACGGCCAC 2490
 QY 661 ACCGGTATTGCCGACTCATGCTATTCTACGCCAGAGCTGACGCCGCTATAGTCGACCG 720
 Db 2491 ACCGGTATTGCCGACTCATGCTATTCTACGCCAGAGCTGACGCCGCTATAGTCGACCG 2550
 QY 721 GCACACGGAACCGGTGAGCTAACCGCGTTTGGTCTATTTCATGGCCGGTTATTGGCT 780
 Db 2551 GCACACGGAACCGGTGAGCTAACCGCGTTTGGTCTATTTCATGGCCGGTTATTGGCT 2610
 QY 781 ATGTGCGGAGGATGATTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAAAACC 840
 Db 2611 ATGTGCGGAGGATGATTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAAAACC 2670
 QY 841 GTTGGTCGGTTTCGATTTTGTAGTGAATAGATCCCAATGATGCCCCACCGGAAGTC 900
 Db 2671 GTTGGTCGGTTTCGATTTTGTAGTGAATAGATCCCAATGATGCCCCACCGGAAGTC 2730
 QY 901 GACCCGGAATCCGGTGAACCTCTCGCTTAAAGCTACGACGCTGCTTTCAAGACCTTACCTA 960
 Db 2731 GACCCGGAATCCGGTGAACCTCTCGCTTAAAGCTACGACGCTGCTTTCAAGACCTTACCTA 2790
 QY 961 AAATACTTCCGATTTCTCACCGGACGGAACCTAAATCACCGGACGCTCGAGATTCAGCTTAT 1020
 Db 2791 AAATACTTCCGATTTCTCACCGGACGGAACCTAAATCACCGGACGCTCGAGATTCAGCTTAT 2850
 QY 1021 CAGCCCAACGATGTCGACGATTTCCGATTTACAGAACTTCGTCGTCGTAACCTGACGAC 1080
 Db 2851 CAGCCCAACGATGTCGACGATTTCCGATTTACAGAACTTCGTCGTCGTAACCTGACGAC 2910
 QY 1081 CAAAGTCGTTTCAAAGCTCCGGAGATGATCCGCGTGCGTCTCCGGTGGTTTACGACAAG 1140
 Db 2911 CAAAGTCGTTTCAAAGCTCCGGAGATGATCCGCGTGCGTCTCCGGTGGTTTACGACAAG 2970

1141 ACAAAGGTCGCAAGATTCGGGATTTTGTAGCAAAATACGCGAAGATTCATCGAACATTAAG 1200
 Db ACAAAGGTCGCAAGATTCGGGATTTTGTAGCAAAATACGCGAAGATTCATCGAACATTAAG 3030
 QY TGGATTGATGCTCCAGATTCCTGCTTCCTCATCTCTGGAACGCTTGGAGAGCCAGAA 1260
 Db TGGATTGATGCTCCAGATTCCTGCTTCCTCATCTCTGGAACGCTTGGAGAGCCAGAA 3090
 QY ACAGATGAAGTCTGATAGGTCCTGTATGATCTCCACCAAGATCAATTTTCAACGAG 1320
 Db ACAGATGAAGTCTGATAGGTCCTGTATGATCTCCACCAAGATCAATTTTCAACGAG 3150
 QY TGTGACGAGAATCTCAAGATGCTGCTGCTGGAACGCTTGGAGAGCCAGAA 1380
 Db TGTGACGAGAATCTCAAGATGCTGCTGCTGGAACGCTTGGAGAGCCAGAA 3210
 QY TCAACTCGCGGTCGATCATCTCCACGAAGATCAACCAAGTCAACCTCGAAGCAGGATG 1440
 Db TCAACTCGCGGTCGATCATCTCCACGAAGATCAACCAAGTCAACCTCGAAGCAGGATG 3270
 QY GTCAACAGAGAATGCTCGGCGGTAAACCAATTCGCTTACTTGTGCTTTAGCCGAGCG 1500
 Db GTCAACAGAGAATGCTCGGCGGTAAACCAATTCGCTTACTTGTGCTTTAGCCGAGCG 3330
 QY TGGCCTAAAGTCTCAGATTCGCTAAAGTTCCTCACTACTGGAAGATTAAGAAACAT 1560
 Db TGGCCTAAAGTCTCAGATTCGCTAAAGTTCCTCACTACTGGAAGATTAAGAAACAT 3390
 QY CTTTACGGCGATAACCGTTACGGAGGAGAGCCTCTGTTCTCCCGGAGAGGAGAG 1620
 Db CTTTACGGCGATAACCGTTACGGAGGAGAGCCTCTGTTCTCCCGGAGAGGAGAG 3450
 QY GAAGACGAAGGATACATCTCTGTTCTGCTTACGACGAGAGACATGGAATCGGAGTTA 1680
 Db GAAGACGAAGGATACATCTCTGTTCTGCTTACGACGAGAGACATGGAATCGGAGTTA 3510
 QY CAGATAGTTAAAGCCTTACGTTAGAGTTGAAGCAACGGTTAACTTCCTCAAGGTT 1740
 Db CAGATAGTTAAAGCCTTACGTTAGAGTTGAAGCAACGGTTAACTTCCTCAAGGTT 3570
 QY CGGTACGGATTCACGGTACATTCATCGGAGGCGGATGTTTCCGGAAGCAGGTCGTGTA 1800
 Db CGGTACGGATTCACGGTACATTCATCGGAGGCGGATGTTTCCGGAAGCAGGTCGTGTA 3630

RESULT 4
 AB028617/c
 LOCUS 52232 bp DNA linear PLN 27-DEC-2000
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MOA2.
 ACCESSION AB028617
 VERSION BA000014
 AB028617.1 GI:5041970
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE
 1 (sites)
 Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
 Structural analysis of Arabidopsis thaliana chromosome 3. I.
 Sequence features of the regions of 4,504,864 bp covered by sixty
 P1 and PAC clones
 DNA Res. 7 (2), 131-135 (2000)
 20277480
 MEDLINE
 PUBMED 10819329
 REFERENCE
 2 (bases 1 to 52232)
 Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
 Direct Submission
 TITLE Submitted (09-JUN-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1532-3, Yata,
 Kisarazu, Chiba 252-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
 Tel:81-438-52-3935, Fax:81-438-52-3934)

COMMENT

Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=MOA2
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremli.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MN21 and the 3' clone is M1E1.

FEATURES

Location/Qualifiers

source

1..52232
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MOA2"
 /clone_11b="Mitsui P1"
 join(1996..2132,2225..2349,2455..2501,2648..2725,
 2807..2931,3021..3087,3216..3278,3370..3439,3529..3629,
 3718..3822,3906..4091)
 /note="gene id:MOA2.2"
 /codon_start=1
 /evidence=not_experimental
 /product="glycolate oxidase"
 /protein_id="BAB01334.1"
 /db_xref="GI:11994212"
 /translations="MEITNVTEYDAIAKQKPKMVDYVYASGAEQDWTQENRNAPAR
 ILFRPAILDVSIVDTMTTTLVGFKISMPIMVAPTAMQKMAHPDEYATARAASAAQT
 MTLSSWATSVSEVASTGPGIRFQYVYVYKRVNVEQLVRAERAGFKAIATLTDTPR
 LGRREDINKRFTLPNLTKNFEGLDKGMDSEANDGLASVAGQIDRTLSWQDVQW
 LGRTKLPLVKGVLTEADARJAIQAGAGIIVSNHGARQLDVYPATISALEBVKVAT
 QGRIPVLDGVRGTDVFKALGASGIFIGRPVVFSLAAGEAGVKVQLQMLRDEF
 ELTVAGCGKSLKEISIRNHITTEWDTFRFSARU"
 complement(join(4591..4659,4763..4924))
 /note="gene id:MOA2.3
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01335.1"
 /db_xref="GI:11994213"
 /translations="MIGRRAGTNRVGMRRDDSLTRFVDSVFYFFRLAEFEILFVLPM
 LITYVIFKDLTSRPEYNIIVEXPGGSDIWP"
 complement(11870..13669)
 /note="gene id:MOA2.4"
 /codon_start=1
 /evidence=not_experimental
 /product="9-cis-epoxycarotenoid dioxygenase"
 /protein_id="BAB01336.1"
 /db_xref="GI:11994214"
 /translations="MASFTATAVSGRWLGGNHTQPLSSQSSDLSSCYSLPMAGRV
 TRKLVNSALHTPPALHFPKQSNPAIVVVKPKASNTKQMLFORAAALDAABG
 FLVSHKHLPLKPTADPSVOIAGNPAVNPQVRRNLVYVGLKPDSTKGYVVRNGAP
 LHPEVCHHFFDGDGMVHVKFHEHGSYACRTQTRFVQERQLGRPVPPKAIHEL
 GHTGIARLMLFYARAAGIYDPAHGTGVANAGLVYFNGRLAMSEDLKYQVQITNG
 DLKTVGDFDGGLESTIAHVPDSEGLFALSVDVSKPILKTKTFRFSDGTGKPD
 VETQDQPTWMDHDFATCFVNVVVDQVFKLPEMIRGSSPVYDKNKAARFGLDKY
 AEISNTKVIDDPAFCFHLWNAWEEPTDEVVIGSCMTTPDSIINESDNLKVLIS
 EIRNLKIGESTRRPIISNDQCNLEAGMVENMLGRKTKFAYLALAPFWPKVSGFA
 KDLITTCGKVKHLYGDNRYGGBELFPGEGGEDEGYILCFVHDEKTXSELQIVNAV
 SLEVAETVKLPSRPVGFHCTFGADDLAKQV"
 join(23980..30324,30421..30528,30616..30660,30841..30913,
 30913..30913)

CDS

CDS

CDS

CDS

```
31021. .31133,31239. .31298,31376. .31492,31588. .31649,
31733. .31793)
/notes="contains similarity to RNA-binding protein
gene_id:"MOA2.5"
/codon_start=1
/evidence=not experimental
/protein_id="BAB01337.1"
/db_xref="GI:11994215"
/translation="MATVTEPMDVAVVDDVKEVSKSEKIIDEGIEKSSIDTSKET
ESLDHKLVMKPKNLPKAKKEFFPYDPKKNQVAKANQILPADDPETTKKSGEE
FDLDKDDORAEHVREVKHMLTGKDAFOAEDILDELOTEALRRVAVAGGLGG
FNSGQGVDCRICDRPHSVLRFAFVADQQAHEALSGGMLGYPVVRVLPSTKAI
LPVNPFLPSREDEMRCTRIYCTINDKKVQADVNFESACGEVETRLRLGGDLH
STRIAFEPALADSALSNLCGMVVGSPQIRVSPSKTPVRPIRPPSTN"
complement(32182. .36456)
/notes="gene_id:"MOA2.6"
/codon_start=1
/evidence=not experimental
/product="disease resistance complex protein"
/protein_id="BAB01338.1"
/db_xref="GI:11994216"
/translation="MANSYLLSSCANVMVERINTSQELVELCKGSSALLKRIKVALV
TANPLADADORAHEVREVKHMLTGKDAFOAEDILDELOTEALRRVAVAGGLGG
LQNLMAGRARAKKIBPKMEKVKVRLLEHVHKEVLEIGLKEYSETREPVQOASRRP
DDUPQGLGRVEDKALNLLSSDDEISIGKPAVSVGMPGVGKTLTEIVENDYR
VTEHFYKMSIAGINFTVTKAVLQDITSSAVNTEDLPSQIKKTLGSKRLL
VLDFWESDESMEFQVAFDAEBSQIVLTRSEIVTAKAEKIQKMLNIBEC
WELTSRPAFNGISVGSINQELGKIAEQCKGLPLAARAIASHLRPNPDWYAV
SKNFSYNTGSLPVKLSDYSLPQLKRCFALCSIFPKGHVFDEELVLLMAIDLLY
OPRSRLEDIGNLDGLVAQFPORLDTWTSFVHMDLMDLAKVASLQDFRLED
DNIPETSTHRSFSDSQDASVAFSICGABFETILPNSPTLSLESLOLTKVLN
PLLNALSGLRILSHVQITNLPKSLGKLKLYLDLSSTKIKELPEFVTCNLQTL
LLSNCRLTSLHRTARIELNLRLLDVLTEPMPGIGIKLLDLSLQKLSNFIQVRLSGA
GLHKLRLSHLGRTRIELSELONAFASEADAGLRKPPFLDGLILKMWKVGSGVPS
GLNKLDELHSLGRTARIELSELONAFASEADAGLRKPPFLDGLILKMWKVGSGVPS
FNALQCGQKELVRLMELPHKLTFCIESYQGGAPPKWLGSFFGTSVTLSSCNLCI
SLPVGQPLSKLISIKRNLQKLVDFGNGNSRGVFPQSLQILKLYGMPRWDEW
ICPELGDIFPCLOKLIIORCPRLKPPGLPSTEVTTSDCLRAVSGNSFRS
LITNPZEPASIPNSRRELSPGNSPKSDASTAOPGFASSOSNDNDEVTSTSLSS
LPKQRTEDPQYITQGLSQPQEEPAVTSARYSGYIDIPSLPSYMSRTSLVDP
KNESILPGSSSYQHQIGIKSPVSPRSRSEALPKSYDDEDEMEVLYKTDLSHME
LPQNLQSLHSDCLGLSPENLTESYFNHLELLIIACHSSFPGGSHPTLTKTYI
RDCKLAFGLSLQTSRYQLEYLFISSCSNLVNFPLSPKLSLSURSDCESFTF
SIHAGLDRLTALSLETRDQCNLETPQGLPTPKLSMLNCKLQALPEKFLG
TSLSLRITLQPEITTPGGFPNRLATCISLQDLTPRIEGLRLLENRLNLDG
QNEIDESPESGLAPKSVFSIRISFENLKTNRKGFHDKAETMBISGCDKLQISI
DEDLPLSLCRLSCLLTETFAEVETEFFKVLNI PYVEIDGEIFS"
36132. .41296
/notes="gene_id:"MOA2.7"
/codon_start=1
/evidence=not experimental
/product="disease resistance complex protein"
/protein_id="BAB01339.1"
/db_xref="GI:11994217"
/translation="WTGIGEMFLAFLQALPQLVSBPSPFRRKRELNENLERLST
ALLTITAVLIDAEKQITNPVEKWNELRDVVHAEDALDDIATEALRLNIGRESS
SNRLQLRGRSLDFDGNSEHLETRLEKVTIRLERLASORNLIGUKELTLPAPOR
LPTTSLVDESEVFGDDDKDIMRFLIPENGKONGITVAIVGIGVGKTLTSLQLYN
DOHVSYGTQKWAHVSSEFVFKITKKVYESVTSRPOCFDLDLVQVKKRLGTG
LPELVLDDLWNEFADLLRQPTTHAAGSQILVTTSRQVASIMCAVHVNLOPL
SDGDWSLPMKTVFCNQPCNLRBEIGDLAERLVHRCGLPLAVKTLGVLRFEGKVI
WVRLSSRIWDLPAKSNLLPVLRVSYVYLPAHLKRCFCAYCSI PKGHAFKDKVIL
WNAEFLQTRDSKLEBELNGEYFSELSRLLQTKTRYIMHDFINELAQFASGES
SKFEDCKLQVSETRVLSYKIAKLPDPDFKNI SHARFLDLSRLELTKLPKSLCYM
QVNSKLLPTLRVLRVLSYKIAKLPDPDFKNI SHARFLDLSRLELTKLPKSLCYM
VNLQTLISYCSLXELPTDINSILNRYLDLIGTKLRQMPRRFRGLKSLQTLTFV
SASDLSRISLGGDLHGLKIKI VELQRVVDAAAEANLNSKKLRARI DFVMTGSS
SENNNPHRTONEAEVTEKPLPHRIEKLAIERVYKGRFPDMLSDPSFSRIVCLR
EQCYCTSLPSLQCLKELHISGMVLRQSLGRKEVFSDDQLRDQDQOPRPSLTLRF
DNLPMQEWELVRVIRGDLFPLSKKLFIIRCEPLTGTLPFLPSLISHIVKGLLDF
QPDHHEYSYRNLQNTLUSIKSSCDTLVKFLPLNHFANUDKLEVDQCSTLSLSLESHLRG
PNALRNLRLNDQNTLQPKNALPQLQVITNCRYLQRPMEQOPQKHHFPHLPS
NVSGSPKSHGSHRSYDSRSSRYD"
42345. .42872
```

CDS

CDS

CDS

CDS

```
/notes="gene_id:"MOA2.8
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB01340.1"
/db_xref="GI:11994218"
/translation="MSDLVIALVAFVAVLFWVFIILSCIMGEXDVHSPPLPBPGLR
MPVYTRKDLGGDTAISVAGTDIYGGSGVGGHGGRRGIITSDTKLPSPFPDRLP
PRPVQRPKSVSRTRKDELIDTGTAMLVASSLLTSGSSASSCSGGSHGCGGGGGGG
GGLGGCGCGGGGGG"
complement(join(42980. .43665,43561. .43668,43952. .44119,
44215. .44460,44539. .44677,44759. .44977,45060. .45754))
/notes="gene_id:"MOA2.9"
/codon_start=1
/evidence=not experimental
/product="terpene synthase-like protein"
/protein_id="BAB01341.1"
/db_xref="GI:11994219"
/translation="MKPKVRDLKSSKNDSNERIRLHLLMLNGIAVHFIEIDEILG
QAFNLDDITAKENDLETTSTMEVPRLGVYMPCKSIGDLSHICHIHYNLAFDPT
VDATNRKGEDGRFKESLAEDIRGMLOLYEAHLGTPSEIDMEALSFTYRLESITS
NHTATSPHUSKHONALYARAHNLEILVAREYISYEQEDHDETLAKFALKNFY
COLHYIDELKDLTKWKBELDASKLPYRDIRIVEYFGALALIFEPYISLGRILVTKI
TMIVTFNDTCDAYTLPESVLSVDSQFQWDLGDIKLPYKIVPGVETLEIEIQ
EMRPGASRIQVAVDEIKKIKAYLAISKWARASHVPTFEYMEFGVQMDHFAAY
SFIAEDCENOTCEWYKSRPKMMEALGVFRINKDINTFEQMSRGEVAKLNCYMK
OHGYSKEAIGEMNKIYSNKYIIMEBYLTTAVPRPILFETDMLNKPILIKITLVM
FRLVNDVGTVEVHLKLTETSRGEVANGLDVYMKQHGVTKEASOEELRKMKNYKV
MEEFMNHHDHPRQVFRCHNIAEIFDVFTVEVDGYCDPKGIENPWSLYLHPIPT
```

Query Match	100.0%	Score	1800;	DB	8;	Length	52232;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	1800;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	ATGGCTTTTTCACGGCAACGGCTCGCTTCTCGGAGATCGCTTGTGGCAATCATACT	60				
DB	13669	ATGGCTTTTTCACGGCAACGGCTCGCTTCTCGGAGATCGCTTGTGGCAATCATACT	13610				
QY	61	CAGCGCCCATATTCGCTTCTTCAAGCTCCGACTTGTAGTTATTTAGCTCTTACCTATG	120				
DB	13609	CAGCGCCCATATTCGCTTCTTCAAGCTCCGACTTGTAGTTATTTAGCTCTTACCTATG	13550				
QY	121	GCAGTCGTGTACAGTAAGTCAATGTTTCATCTCGCTTCAACCTCCAGCTCTT	180				
DB	13549	GCAGTCGTGTACAGTAAGTCAATGTTTCATCTCGCTTCAACCTCCAGCTCTT	13490				
QY	181	CATTTCCCTTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCAAGCCAAAGAA	240				
DB	13489	CATTTCCCTTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCAAGCCAAAGAA	13430				
QY	241	TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGGGCGGCGAGCTTGACCGGGCG	300				
DB	13429	TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGGGCGGCGAGCTTGACCGGGCG	13370				
QY	301	GAGGGTTTCTTGTGAGCCACAGAGAGCTACACCGCTTCTTAAACCGCTGATCCTAGT	360				
DB	13369	GAGGGTTTCTTGTGAGCCACAGAGAGCTACACCGCTTCTTAAACCGCTGATCCTAGT	13310				
QY	361	GTTCAATCGCCGAAATTTTCTCGGTGAATGAACACCGCTCGGGGTAACTCTTCG	420				
DB	13309	GTTCAATCGCCGAAATTTTCTCGGTGAATGAACACCGCTCGGGGTAACTCTTCG	13250				
QY	421	GTGCTCGAAACCTTCCCGATCCATCAAGAGATGTATGTGCGCAACGAGACTAACCCA	480				
DB	13249	GTGCTCGAAACCTTCCCGATCCATCAAGAGATGTATGTGCGCAACGAGACTAACCCA	13190				
QY	481	CTTCAGAGCGGTGACAGGTACACCTTCTTCGACGAGAGCGTATGTTTACCGCGTC	540				
DB	13189	CTTCAGAGCGGTGACAGGTACACCTTCTTCGACGAGAGCGTATGTTTACCGCGTC	13130				
QY	541	AAATTGGAACACCGTTTCACTAGTACGCTTGGCGGTGTTTACTCAGACTAACCGGTTGTT	600				
DB	13129	AAATTGGAACACCGTTTCACTAGTACGCTTGGCGGTGTTTACTCAGACTAACCGGTTGTT	13070				

QY 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGGCAATCGGTGACGTTCAAGGCGAC 660
Db 13069 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGGCAATCGGTGACGTTCAAGGCGAC 13010
QY 661 ACCGGTATTGCCCGACTCATGTTATTTACGCGACAGCTGACGCGGTATAGTCGACCGG 720
Db 13009 ACCGGTATTGCCCGACTCATGTTATTTACGCGACAGCTGACGCGGTATAGTCGACCGG 12950
QY 721 GCACACGGAACCGGTGTAGTCAACGCGGTTTGGTCTATTTCAATGGCGGTTATTTGGCT 780
Db 12949 GCACACGGAACCGGTGTAGTCAACGCGGTTTGGTCTATTTCAATGGCGGTTATTTGGCT 12890
QY 781 ATGTCCGAGGATGATTATCTTACCAAGTTACAGTTCACCTCCCAATGGAGATTTAAACCC 840
Db 12889 ATGTCCGAGGATGATTATCTTACCAAGTTACAGTTCACCTCCCAATGGAGATTTAAACCC 12830
QY 841 GTTGTGCGTTCGATTTTATGATGGACAAATAGAAATCACAATGATTGCCACCCGGAAGTC 900
Db 12829 GTTGTGCGTTCGATTTTATGATGGACAAATAGAAATCACAATGATTGCCACCCGGAAGTC 12770
QY 901 GACCCGGAATCCGGTGAACCTTCGCTTTAAGCTAGCAGCTGCTTTCAAAGCCTTACCTA 960
Db 12769 GACCCGGAATCCGGTGAACCTTCGCTTTAAGCTAGCAGCTGCTTTCAAAGCCTTACCTA 12710
QY 961 AAATACCTCCGATTTCTACCGGACGGAACATAATCACCGGAGCTCGAGATTCAGCTTGAT 1020
Db 12709 AAATACCTCCGATTTCTACCGGACGGAACATAATCACCGGAGCTCGAGATTCAGCTTGAT 12650
QY 1021 CAGCCACGATGATGACAGATTTCCGATTTACAGAACTTCGTCGTCTACTCTGACGAC 1080
Db 12649 CAGCCACGATGATGACAGATTTCCGATTTACAGAACTTCGTCGTCTACTCTGACGAC 12590
QY 1081 CAACTGCTTTTCAAGTCCGCGAGATGATCCGCGTGGGTCTCCGTTGTTTACGACAG 1140
Db 12589 CAACTGCTTTTCAAGTCCGCGAGATGATCCGCGTGGGTCTCCGTTGTTTACGACAG 12530
QY 1141 AACAGGTCCAGATTCGGATTTAGACAAATACCGCGAGATTCATCGAACATTAAG 1200
Db 12529 AACAGGTCCAGATTCGGATTTAGACAAATACCGCGAGATTCATCGAACATTAAG 12470
QY 1201 TGGATTGATCTCCAGATTCCTTCTGCTTCCATCTCTGAAACGCTTGGGAAGGCGAGAA 1260
Db 12469 TGGATTGATCTCCAGATTCCTTCTGCTTCCATCTCTGAAACGCTTGGGAAGGCGAGAA 12410
QY 1261 ACAGATGAAGTCTGCTGATAGGCTCTGATGACTTCCACGACACTCAATTTTCAAGGAG 1320
Db 12409 ACAGATGAAGTCTGCTGATAGGCTCTGATGACTTCCACGACACTCAATTTTCAAGGAG 12350
QY 1321 TCTGACGAGAACTCTCAAGAGTGTCTGCTGTAATCGCTGAATCTCAAAACCGGTGAA 1380
Db 12349 TCTGACGAGAACTCTCAAGAGTGTCTGCTGTAATCGCTGAATCTCAAAACCGGTGAA 12290
QY 1381 TCAACTCCGCTCGGATCATCTCCAGAAAGATCAACAGTCAACCTCGAAGCAGGATG 1440
Db 12289 TCAACTCCGCTCGGATCATCTCCAGAAAGATCAACAGTCAACCTCGAAGCAGGATG 12230
QY 1441 GTCAACGAGAACTGCTCGCGCTTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCGG 1500
Db 12229 GTCAACGAGAACTGCTCGCGCTTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCGG 12170
QY 1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTCGAGAAGTTAAGAACAAT 1560
Db 12169 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTCGAGAAGTTAAGAACAAT 12110
QY 1561 CTTTACCGGATACCGGTACGAGGAGAGGCTTCTGTTTCCCGGAGGAGGAGGAG 1620
Db 12109 CTTTACCGGATACCGGTACGAGGAGAGGCTTCTGTTTCCCGGAGGAGGAGGAGGAG 12050
QY 1621 GAAGACGAAGGATACATCTCTGTTTGGTTTCCAGCAGGAGAGACATCGAAATCGGAGTTA 1680
Db 12049 GAAGACGAAGGATACATCTCTGTTTGGTTTCCAGCAGGAGAGACATCGAAATCGGAGTTA 11990

QY 1681 CAGATAGTTAAACCGCTTTAGCTTAGAGTTTGAAGCAACGGTTAAACTTCGCTCAAGGTT 1740
Db 11989 CAGATAGTTAAACCGCTTTAGCTTAGAGTTTGAAGCAACGGTTAAACTTCGCTCAAGGTT 11930
QY 1741 CCGTACGGATTTACGGTACATTCATCGGAGCGGATGATTTGGCGAAGCAGGTCGTGTGA 1800
Db 11929 CCGTACGGATTTACGGTACATTCATCGGAGCGGATGATTTGGCGAAGCAGGTCGTGTGA 11870

RESULT 5
AY056255 2331 bp mRNA linear PLN 18-SEP-2002
LOCUS Arabidopsis thaliana putative 9-cis-epoxycarotenoid dioxygenase
DEFINITION (At3g14440) mRNA, complete cds.

ACCESSION AY056255.1 GI:15810432
VERSION 1
KEYWORDS Full cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 2331)
Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K., Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R. and Theologis A.
Arabidopsis Full Length cDNA Clones

Unpublished
2 (bases 1 to 2331)
Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C.C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R. and Theologis A.
Direct Submission

TITLE Submitted (12-SEP-2001) Plant Gene Expression Center, 900 Buchanan Street, Albany, CA 94710, USA
JOURNAL RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Yamada, K., Banh, J., Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES
Location/Qualifiers
1. 2331
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAPL08-11-H16 (R11094)"
/note="This clone is in a modified pBluescript vector source

QY 1501 TGGCTAAAGTCTCAGGATTCCTAAAGTTGATCTCACTACCTGGAGAGTTAAGAACAT 1560
Db 1623 TGGCTAAAGTCTCAGGATTCCTAAAGTTGATCTCACTACCTGGAGAGTTAAGAACAT 1682
QY 1561 CTTTACGGCGATAACCGTTACGGAGAGAGCTCTGTTTCTCCCGGAGAGAGGAGAG 1620
Db 1683 CTTTACGGCGATAACCGTTACGGAGAGAGCTCTGTTTCTCCCGGAGAGAGGAGAG 1742
QY 1621 GAAGACGAAGGATACATCTCTGTTTCTGTTTACGACGAGAGACATGGAATCGGAGTTA 1680
Db 1743 GAAGACGAAGGATACATCTCTGTTTCTGTTTACGACGAGAGACATGGAATCGGAGTTA 1802
QY 1681 CAGATAGTTAAACCGGTTAGCTTACGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1740
Db 1803 CAGATAGTTAAACCGGTTAGCTTACGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1862
QY 1741 CCGTACGGATTTACGGGTACATTCATCGGAGCGGATGATTTGGCGAAGCAGGTCTGTGA 1800
Db 1863 CCGTACGGATTTACGGGTACATTCATCGGAGCGGATGATTTGGCGAAGCAGGTCTGTGA 1922

RESULT 6
AX652128 95769 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 1979 from Patent WO03000898.
DEFINITION AX652128
ACCESSION AX652128
VERSION AX652128.1 GI:29154942
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katzirir, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, F. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 1979 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
1. 95769
Source /organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

BASE COUNT 30958 a 16961 c 16682 g 31168 t
ORIGIN

Query Match 47.9%; Score 862.8; DB 6; Length 95769;
Best Local Similarity 73.1%; Pred. No. 6.2e-248;
Matches 1139; Conservative 0; Mismatches 407; Indels 12; Gaps 2;

QY 245 ACATTAACAGATGAATTTGTTTCCAGAGAGCGCGCGCGACGTTGGACGCGCGAGG 304
Db 33987 AGAATCCAAAATTAACCGGTTGACAGAAACCGCGCTATGTTTACGCGGTGAAA 34046
QY 305 GTTTCCTGTCACGACGAG---AGCTACACCGCTTCTTAAACGGCTGATCTAGTG 361
Db 34047 ACCTCATGATCTCACAGAGCGCGCGCTATCCCATCTTAAACGGCGATCTCGCG 34106
QY 362 TTCAGATCGCGGAAATTTGTTCCGTTGAATGAACAGCGCGCGCGCTAATCTTCGG 421
Db 34107 TCCAAATAGCGGAACTTCTCCCGGTACCGGAGAACTGTGTCGATACCTTCGG 34166
QY 422 TGTCGCGAAAACCTTCCGATTCATCAAGAGAGTGTATGTGCGCAACGGAGTACCCAC 481
Db 34167 TGACTCGAAACAGTACAGAAATGCAATCAAGAGAGTTTACGTCAAGAACCGAGAAATCCAC 34226
QY 482 TTCAGAGCGCGGTGACAGGTACCACTTCTTCGACGAGACGATGTTTACGCGGTCA 541
Db 34227 TTCACAAACAGTCTCCGCCACCAATTTGTTCAAGAGAGCGGTATGTTACAGTCTC 34286
QY 542 AATTCGAAACAGGTTACGATGCTACGCTTCCGCGTTTACTCAGACTAACCGGTTTGTTC 601

Db 34287 GGTTCGATACCGCTCGGTTAGCTACGTTGCGTTTACGAAACAAACCGTTGTTTC 34346
QY 602 AGGAACGTCATTAATGGTTCGACCGGTTTCCCAAAGCCATCGGTGAGCTTCACGGCCAC 661
Db 34347 AAGAACGTCATTAATGGTTCGTCGCGGTTTCCCAAAGCAATTCGAGAGCTTCACGGACAT 34406
QY 662 CCGGTATTCGCCAGCTCATCTATCTACGCGAGAGCTCGACCGGTATAGTCGACCCGG 721
Db 34407 TAGGAATCGCGAAGCTTATGCTCTTCAATACCGCGGGCTATTGCGGTATAGTCGACCCGA 34466
QY 722 CACACGGAACCGGTGAGCTTAACCGCGGTTTGGTCTTATTTCAATGCGCGGTTATTCGGTAT 781
Db 34467 CCGGAGGACTCGGTGTCGCTTAACCGCGGTTTATTTTCAATGCTCATCTCTTAGCCA 34526
QY 782 TGTCGGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGAGATTAATAACCG 841
Db 34527 TGTCGGAGGATGATTTACCGTACCATGTCAAAGTTACTCAAACCGGAGATTTAGAAACTT 34586
QY 842 TTGCTCGGTTTCGATTTGATGGAACAATTAGAATCCCAATGATGTCGCCACCGAAAGTCG 901
Db 34587 CCGGTTCGATCAATTTTCGACGGTCACTTAAATCAACAATGATAGCCACCGGAAATCG 34646
QY 902 ACCCGGAATCCGGTGAATCTTTCGCTTTTAAGCTACGAGCTGTTTCAAGCCCTTACTAA 961
Db 34647 ATCCGGAACCGGAGAACTCTTCGCTTTTAAGCTACGAGCTGTTTCAAGCCCTTACTAA 34706
QY 962 AATACTTCGATTTCTACCGGACGGAATAAATCAGCGAGCTCGAGATTCAGCTTGATC 1021
Db 34707 AATACTTCGATTTCAATCCGACGGTGAAGAAATCAGAGCTCGAGATTCGCTTGATC 34766
QY 1022 AGCCACGATGATGACAGATTTCCGATTTACAGAGAACTTCGTCGTCGTCGTCGTCGAC 1081
Db 34767 AACCACGATGATCCAGATTTTCGCGATCACTGAGAAATTCGTCGTCGTCGTCGTCGAC 34826
QY 1082 AAGTCGTTTCAAGCTGCGGAGATGATCGCGGTGCTCTCCGTTGGTTTACGACAAGA 1141
Db 34827 AAGTCGTTTCCGATTTCCGAGATGATAGAGAGTGTCTCCGTTGGTTTACGACAGA 34886
QY 1142 ACAAGTTCGCAAGATTCGGATTTTAGACAAATACGCGGAGATTCATCGAACTTAAGT 1201
Db 34887 AGAAGAAATCAAGATTCGGATTTTGAATAAAACGCGAAGATGCTTCTGTCGATCAAT 34946
QY 1202 GGATGATGTCGAGATTTGCTTCTGCTTCATCTCTGGAACGCTTCGGAGAGCCAGAAA 1261
Db 34947 GGATCGAAGTACGAGATTTGTTTGTTCATCTATGGAATCTTCGGAAGAACAGAAA 35006
QY 1262 CAGATGAAGTCTGTCGATAGGCTCTGTATGATCTCCACGAGCTCAATTTTCAACGAGT 1321
Db 35007 CAGACGAGGTTGTCGTCGATTCGATCATGATGACGCCACCTGATTCATTTTCAACGAA 35066
QY 1322 CTGACGAGATCTCAAGAGTGTCTGTCTGAATTCGCTGATCTCAAAACCGGTGAAT 1381
Db 35067 ACAGCGAAACACTTCAGAGTGTTCGTCGAGATAAGGCTTAAACCTTAAACAGGGGAT 35126
QY 1382 CAATCCCGCTCCGATCATCTCCACGAGATCAACAGTCAACCTCGAGAGGAGTGG 1441
Db 35127 CAACAGTACCGGTTATCTC-----GGACAAGTTAATCTCGAAGCTGGTATGG 35177
QY 1442 TCACAGAAACATGCTCGCGCTTAAACCAAATTCGTTACTTGGCTTTAGCCGAGCCGT 1501
Db 35178 TCACCGGAAATTTATAGGTAGAAAACCGCGTATCTTATCTGCTTTTAAACGAAACCGT 35237
QY 1502 GGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTCGGAGAGTTTAAAGAACATC 1561
Db 35238 GGCTTAAAGTCTCGGTTTCGCTTAAAGTGAATTCGACCGGAGAGATTCGAAGATATA 35297
QY 1562 TTTCAGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAAAGGAGGAGG 1621
Db 35298 TTTCAGGAGAGGAGAAATACGAGGAGAGCTCTGTTTCTTACCTTCGCTTCGAGCGGAG 35357
QY 1622 AAGACGAGGATACATCTCTGTTTCTGTTTCAACGACGAGAGATGATAATCGAGTTAC 1681

Db 35358 AAGACGAGGTTACATAATGTTGTTCTTCAAGCAGGAGGAGGTGAATCGGAACATCC 35417
LOCUS
Qy 1682 AGATAGTTAACGCGGTTAGCTTAGAGTTGAACAAACGGTTAAACTTCGTCGAAGGTTTC 1741
Db 35418 AACTCATAACGCGGTTAATATGAAGCTTGAAGCTTACCGTACGCTTCGTCGAGAGTGC 35477
Qy 1742 CGTACGAGTTACGCTACATTCATCGAGCCCATGATTTGGCAGACGAGTCGTTG 1799
Db 35478 CTTACGCTTTCCACGGAACGCTTCATCAGTAAGGAAGATTATTCGAAGCAAGCTTTGTG 35535

RESULT 7
AC013430 95769 bp DNA linear PLN 26-MAY-2000
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome I, complete sequence.
ACCESSION AC013430
VERSION AC013430.5 GI:8096768
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 95769)
AUTHORS Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C., Shinn,P., Altafi,H., Bei,Q., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome I
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 95769)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 95769)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 95769)
AUTHORS Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE 5 (bases 1 to 95769)
AUTHORS Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On May 26, 2000 this sequence version replaced gi:6921155.
FEATURES Location/Qualifiers

source
1. 95769
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F3f9"
join(71..402,574..649,733..933)
/note="similar to oxidosqualene cyclase dbj|BAA33462.1"
/codon_start=1
/evidence=not experimental
/product="F3f9.1"
/protein_id="AAF71792.1"
/db_xref="GI:8052528"
/translation="MGHVQLTFCYFFPFFSFFLLFFLNAXNFIILISVDLTGVV
FELEKVEQVIPPVVEDGSETHETRRMYGEEFSSRLCRPMTVGLKSLDLSS
FLHWYSSFLHLEIFDAHREILRYVYHLNDGGWGLHVEGKSFMCFTALNYI
CLRLREGDEGREYACKLTKMPWVTFLLIYNNSCFPSI"
complement(join(3438..3746,3868..4032,4110..4664))
/note="similar to F-box domain gb|AAF14684.1"
/codon_start=1
/evidence=not experimental
/product="F3f9.2"
/protein_id="AAF71813.1"
/db_xref="GI:8052549"
/translation="MLASWISQLTSMKHTKNLILANGRFAPTEGDNVILKLTETN
HPSRTLSLYIQLDAQFNCKKYLDFQFAEVDVNFVEISSCPSLEVLILQI
IFFPNSGLKTDHKTILSMSCNQIDSIIEVRAARLDILSIEYIPCSDNVLEIPRL
QFGNYYVAGRLPLHTSINISCPQKQESNGWKMDTYATSPASLSLDLKPREV
EVLAKILAVTEKMEIVEISFKNPNLGEENSDGGAQNNLWEKAEFPFNADPRVT
IWMNFKGSNKEQFALASRFVMQKTVMKWKMTKTSFDEKKKETEAAVAKLKLPGKN
BELSIECF"
join(7138..7400,7491..7887)
/note="unknown protein; similar to EST gb|AI999747.1"
/codon_start=1
/evidence=not experimental
/product="F3f9.3"
/protein_id="AAF71793.1"
/db_xref="GI:8052529"
/translation="MERTGLNLLKLSLVVVVGSYACAPAPWNPNGSRPGTCDHVE
CPTKLVAGYGFIRDMALMISTFISLSTQATKTFRPLNRYIEGDNKSNVK
NMTPAVTAQATPGRSVTVYLPKKNQNPPOADDLHVRSRTPTVAVRQIGGVYS
NNVADAAALMESLRDSNWLPIETKSGKLPAFLAVYINPPSHTTARVINEINWPN
M"
join(8524..8759,8846..9287)
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="F3f9.4"
/protein_id="AAF71794.1"
/db_xref="GI:8052530"
/translation="MEALSIFKLSFLITLLSGGSLDLSLPPNCRNACPSYEVVHA
GNGYRHSYNTVMTSTPIODISLNEASGNGWQLSDYMGNNNDYHORIAPLYIT
QVSNLSYFVSEFVFKAFQBDPPGNLHVQRWDSRYAVKQISGVADHKIGQVA
ELKSLQGTVMKALEKREKRGVGSWAYVVAQSFQWQSVNEIWFPEMEDEE
TVSIVTQ"
join(13444..13814,13956..14310,14399..14662)
/note="similar to dioxygenase protein"
/codon_start=1
/evidence=not experimental
/product="F3f9.5"
/protein_id="AAF71795.1"
/db_xref="GI:8052531"
/translation="MVLKSPVAIPKSGFSIPIVDMSDPSKHALVKACDFGFFKV
INHGVSALVSEIHEVDFSLPKSEKTQVAGYFGNGSKIGNGVDGWWVEYLLMN
ANHDSGSLPFLSKPGTFRNALLEYTSVRKMTFDVLEKIDGLGKIPNTLSKL
VSDQNTSILRLNHVPCPLGNKNTGKNGVIGFGEHTDPOIISVLRGNTSGQLNL
NDGSLVPPDHTSFEFVNGSLQWNTNGRFKSVHRVLANCKSRVSMVYFAGPSLT
QRIAPLTLCDNEDELVEETWSEYKNTSNRSLDNRLOQFERKTIKNLLN"
complement(join(15755..16715,16810..16829))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="F3f9.6"

Db	35067	ACGACGAAACACTTCAGAGTGTTTGTTCGAGAGTAAGGCTAAACCTTTAAACACGGGAAT	35126
QY	1382	CAACTCGCGTCGGATCATCTCCAACGAGATCAACAAAGTCAACCTCGAAGCAGGGATGG	1441
Db	35127	CAACAGTAGACGGGTTAICTC-----GGAACAAAGTTAACTCTCGAAGCTGGTATGG	35177
QY	1442	TCAACAGAAACAATGCTCGGCCGTAAACCAAAATTCGCTTACTTTGGCTTTAGCCGACCGT	1501
Db	35178	TCAACCGGAATTTATTAGTAGTAAACACCGGTATGCTTATCTGGCTTTTAAACCGAACCGT	35237
QY	1502	GGCTAAAGTCTCAGGATTCGGTTAAAGTTGATCTCACTACTGGAGAAAGTTAAGAAACATC	1561
Db	35238	GGCTAAAGTGTCCGGTTTCGCTAAAGTGACATTTCGACGAGAGATTTCGAAATATA	35297
QY	1562	TTTACGGCGATAACCGTTACGAGGAGAGCCCTCTGTTCTCCCGGAGAGAGGAGAGG	1621
Db	35298	TTTACGGAGAGGGGAAATACGGAGGAGAGCCCTCTGTTCTACCTCTCCGGTGACGGAGAAG	35357
QY	1622	AAGACGAAGGATACCTCTCTGTTTCAGACGAGAGACATGGAATTCGGAGTTAC	1681
Db	35358	AAGACGGAGGTTACATAATGGTGTTCGTTTCACGACGAGAGAGTGGAATTCGGAACATCC	35417
QY	1682	AGATAGTTAACCGCGTTAGCTTAGAGGTTGAAGCAACGGTTTAAACTTCGCTCAAGGGTTC	1741
Db	35418	AATCTATAACCGCGTTAATGAAGCTTGAAGCTACCGTAACGCTTCGCTCGAGAGTGC	35477
QY	1742	CGTACGGATTCACGCTACATTCATCGAGCCCGATGATTTGGCGAAGCAAGGTCGTGTG	1799
Db	35478	CTTACCGGTTTCCACGGAACGTTTCATCAGTAAGGAAGATTATTCGAAGCAAGCTTTGTG	35535

RESULT	8
AXI48316	
LOCUS	1818 bp DNA linear PAT 08-JUN-2001
DEFINITION	Sequence 15 from Patent EP1116794.
ACCESSION	AXI48316
VERSION	AXI48316.1 GI:14347203
KEYWORDS	Lycopersicon esculentum (tomato)
SOURCE	Lycopersicon esculentum
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asterales; Lamiales; Solanales; Solanaceae; Solanum; Lycopersicon. 1 Tsuchi, S., Kobayashi, M. and Shinozaki, K. Transgenic plants carrying neoxanthin cleavage enzyme gene Patent: EP 1116794-A 15 18-JUL-2001; Riken (JP)
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

```

1. 1818
/db_xref="Caioni:1861"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41199.1"
/db_xref="GI:14347204"
/translation="MATTSHANTWIKTLKSLMPSSKEFGFASNSTLLKNOHQCSL
NINSSLAQPIILHFFPKOSSNYQFPKNTISHPKQENNSSSSSTKNNLVOKAAAMAL
DAVESALTQHELEHLPEDPADRPVQSGNFAPENPVQSLPVTGKIKPCVQGVYVR
NGANPLPEPTAGHLLFPDGDWGHAVQPKNGSAYACRTETERLVOKLAMEDLPVHPKA
IGELHSGISGARLMLFFYARGLFGLVHDSKGTGVANAGLVYFNRLVKAMEDDLPPVHK
VPTPGDLKTEGEPFDQGLATMIAPHKLPDVSQELFALSVDYIQTPYLKPYKFRKSG
EKSNDVLPVEDPTMHDDFAITENFVYPDQVVFKNSEMIKGGSVYDQKQKVRFG
ILDXAYDGDSDLKWEVDPDCCFCFELHWNABEATDEIVTIGSCWTTPDPSIPNECDGGL
KSVLEARNLTKTKGSTRKIIENPDQVNLGAPNRNKLGRKTYAVLAIAEPMPK
VSGFAKNVLFGEVEKTYTGDKYKGEGEPLFPDQPNKSGEEDDGYIIAFVHDEKWKSE
LQIVNWSLKLEATVPLKPSRVPYFGHTFINANDLANQA"
569 a 383 c 369 g 497 t

```

BASE COUNT	569 a	383 c	369 g	497 t
ORIGIN				
Query Match		45.8%	Score 824.6;	DB 6; Length 1818;

	Best Local Similarity	69.6%;	Pred.No. 8.4e-237;		Mismatches	0;	Indels	6;	Gaps	1;
	Matches	1134;	Conservative							
QY	170	CTCCAGCTCTCAATTCCTTAAGCAATCATCAAACTCTCCGCCATTTGTTTAAGCCCA	229							
DB	185	CTTCAAAATATCAAAACACAAGATAATACAATTCACCCAAAACAAGAAACAACA	244							
QY	230	AAGCCAAAGAATCCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCAGCGT	289							
DB	245	ACTCTCTCTCTTCACTTCCAACTTCCAACTGGAATTTAGTCAGAAAGCAGCAATGGCTT	304							
QY	290	TGGACGGCGGAGGGTTTCCCTTGTCAGCAGCAGAGAAGCTACACCGCTTCTTAAACGG	349							
DB	305	TAGATGCTGTAGAAAGTGCTTTAACTAAACATGAACCTTGAACACCCCTTTCGGAACAAG	364							
QY	350	CTGATCTAGTGTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCCGTCCGGC	409							
DB	365	CCGACCCACAGTCCAGATTTCTGGAAATTTTGTCTCGGTACCGGAAATCCAGTCTGTC	424							
QY	410	GTAATCTTCGGTGGTCGGAAACTTCCGATTCCATCAAGGATGTATGTGCGCAAGC	469							
DB	425	AATCTCTTCGGTCCCGGAAATATCCCAATGTGTTCAAGGCGCTTTACGTTTCGAAACG	484							
QY	470	GAGCTAACCCACTTCACGAGCCGGTGAACAGGTCACCACTTCTTCACGAGAGCGGTATGG	529							
DB	485	GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCAATTTCTTCAGCGCAGCGTATGG	544							
QY	530	TTACGCGCTCAAAATTCGAACACGGTTACGCTAGCTAGCTATGCGGTTTACTCAGACTA	589							
DB	545	TTACGCGCTTCAATTCAAAAATGGTGGCTAGTACGTTTGGCGTTTCTCTGAAACAG	604							
QY	590	ACCGGTTGTTTCAGGAACGTCAAATGGGTGACCGGTTTTCCTCCAAAGCCATCGGTGAC	649							
DB	605	AGAGGCTTGTTCAGAAAGAGCTTTGGGTGCGCCCTTTTCCCTAAAGCCATTTGTTGAAT	664							
QY	650	TTCAAGCGCACACCGGTATTTGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCGGTA	709							
DB	665	TACATGGTCACCTCGAAATGCAAGGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGAC	724							
QY	710	TAGTCAGCCGGCACACGGAAACCGGTGTAGCTAAACGCCGGTTTGGTCTATTTCAATGGCC	769							
DB	725	TTGTTGATCACAGTAAGGAACCTGGTGTTCGAAAGCGCGGTTTAGTCTATTTCAATAACC	784							
QY	770	GGTTATTTGGGATGTGCGAGGATGATTACCTTACCAAGTTACAGTACACTCCCAATGGAG	829							
DB	785	GATTACTGTGATGTCTGAAGATGATTGCGCTTACCATGAAGGTAACACCCACCGCG	844							
QY	830	ATTTAAAAACCGTTGTCGGTTTCGATTTTGATGGACAAATAGAATCCCAATGATTGCC	889							
DB	845	ATCTTAAAAACAGAGGTCGATTCGATTTTCGAGGCCAGCTAAAAATCCCATGATAGCTC	904							
QY	890	ACCGGAAGTCGACCCGGAAATCCGGTGAATCTTTCGCTTTAAGCTACGAGCTGTTCAA	949							
DB	905	ACCCAAAGCTCGACCCCAATATGTCGAGTATTTGCTCTTAGCTACGATGTGATTCAGA	964							
QY	950	AGCCTTACCTTAAATACTTCCGATTTCTACCGGACGGAATTAATCACCGGAGCTCGAGA	1009							
DB	965	AGCATACCTCACTACTTCAGATTTTCAAAAAATGGGAAAAATCAATGATGTTGAAA	1024							
QY	1010	TTAGCTTGATTCAGCCAAACGATGACGATTTCCGATTACAGAGAACTTCGTGCTCG	1069							
DB	1025	TTCCAGTTGAAGCCCAACAAATATGTCATGATTTGCAATTTACTGAGAACTTCGTGCTCA	1084							
QY	1070	TACCTGACCAAGAGTCGTTTTTCAAGCTGCCGAGATGATCCCGGTGGGTCTCCGCTGG	1129							
DB	1085	TTCTTGATCAACAGTCGTTTTCAAGATGCTGAATGATCCGTGGAGTTTACCCTGGTGG	1144							
QY	1130	TTTACGACAGAAACAAGTTCGAAGATTCGGGATTTTAGCAAAATACCGCGAGATTCAT	1189							
DB	1145	TTTACGACAGAAACAAAGTTTCCCGATTTCCGATTTCCGATTTTCGATAATTCGGAAGAT	1204							
QY	1190	CGAATTTAAGTGATTGATGCTCCAGATTTGCTTCTCTTCCATCTCTGGAACGCTTGGG	1249							

Db 1205 CTGATTGAAATGGGTGAAGTACCTGATTTGTTCTGTTTCCACCTCTGGAATGCTTGGG 1264
Qy 1250 AAGAGCCAGAAACAGATGAAGTCGTGATAGAGGTCCTGTATGACTCCACGAGACTCAA 1309
Db 1265 AAGAGCAGAAACAGATGAATCGTTGATTTGTTGTTGATGATGACACCAAGACTCCA 1324
Qy 1310 TTTTCAAGAGTCTCAGAGAAATCTCAGAGTGTCTGTCTGAAATCCGCTGAACTCTCA 1369
Db 1325 TTTTCAATGAATGTGATGAAGGGCTAAAGAGTGTCTTATCCGAAATCCGCTCAATTGA 1384
Qy 1370 AAACCGGTGAATCAACTCCGCTCGATCATCTCCAAAGAGATCAACAAAGTCAACCTCG 1429
Db 1385 AAACAGGGNAATCAACAGAAATCCATAATCGAAACCCGGATGAACAAAGTGAATTAG 1444
Qy 1430 AAGAGGGATGTCAACAGAAATCTGCTGGCCGCTGAAACCAAAATTCGTTACTTGGCTT 1489
Db 1445 AAGCTGGAATGTGAACCGAAACAACTCGGAAGGAAACAGAGTATGCTTATTGCGTA 1504
Qy 1490 TAGCCGAGCGGTGGCTAAAGTCTCAGGATTCGCTTAAAGTTCATCTACTACTCTGAGAG 1549
Db 1505 TCGCTGAACCATGGCCAAAGTTCTGTTTGGAAAGTAAACCTGTTACCGGTGAAG 1564
Qy 1550 TTAAGAAACATCTTTACGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAG 1609
Db 1565 TTGAGAAATTCATTTATGTTGACAAACAAATATGTTGGGGAACCTCTTTTATCCAGAG 1624
Qy 1610 -----AAGGAGGAGAGAGAGAGGATACATCTCTGTTTCTGTTTCTCAGCAGAGAGA 1663
Db 1625 ACCCAACAGCAAGAGAGAGAGAGTGTATATTTTAGCTTCTGTTTCCATGAGAAAG 1684
Qy 1664 CATGAAATCGGAGTTACAGATAGTTAAGCCGTTAGCTTAGAGTTTGAAGCAACGTTA 1723
Db 1685 AATGAAATCAGAACTGCAAAATGTTTAAAGCAATGAGTTTGAAGTTGAGGCAACTGTGA 1744
Qy 1724 AATCTTCGTCAGGTTCCGTACGATTTACGATTTACGATTTACGAGCCGATTTGG 1783
Db 1745 AGCTTCCATCAGAGTTCCTTATGATTTTATGATTTTATGAAACATTCATAAACCCCAATGTTGG 1804
Qy 1784 CGAAGCAGG 1792
Db 1805 CAAATCAGG 1813

BD017436 1818 bp DNA linear PAT 27-AUG-2002
Transgenic plant using neozanthine cleaving enzyme gene.
BD017436
BD017436 1 GI:22558612
JP 2001258579-A/8.
Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 1818)
Luchi,K.; Kobayashi,M. and Shinozaki,K.
Transgenic plant using neozanthine cleaving enzyme gene
Patent: JP 2001258579-A 8 25-SEP-2001;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
OS Lycopersicon esculentum (tomato)
PN JP 2001258579-A/8
PD 25-SEP-2001
PF 11-JAN-2001 JP 2001003476
PI KIYOSHI TUCHI,MASATOMO KOBAYASHI,KAZUO SHINOZAKI PC
C12N15/09,A01H5/00,C12N9/10,C12N9/02,C12N15/00,C12N5/00 CC
Transgenic plant using neozanthine cleaving enzyme gene PH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..1818
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"

BASE COUNT 569 a 383 c 369 g 497 t
ORIGIN
Query Match 45.8%; Score 824.6; DB 6; Length 1818;
Best Local Similarity 69.6%; Pred. No. 8.4e-237;
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;
Qy 170 CTCAGCTCTTCAATTTCCCTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCA 229
Db 185 CTTCAAATATCAACACCAAGATAATACAAATTTTCAACCCAAACCAAGAAACAACA 244
Qy 230 AAGCCAAAGATCCACACATCAACATGAATTTGTTCCAGAGAGCGGCGGAGCGT 289
Db 245 ACTCTCTCTTCTTCAACTTCCAACTTCCAACTTGAATTTAGTCAGAAAGCAGCAATGGCTT 304
Qy 290 TGGACCGCGCGAGGTTTCTTTGTTCAGCCACGAGAAAGTACACCCGCTTCTCTTAAACCG 349
Db 305 TAGATGCTGTAGAAAGTGTCTTTAACTTAAACATGAACCTTGAACACACCTTTGCCGAAACAG 364
Qy 350 CTGATCTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAATGAACAGCCGCTCCGCG 409
Db 365 CCAGCCACGAGTCCAGATTTCTGGGAATTTTGTCTCCGTACCGGAAATTCAGTCTGTC 424
Qy 410 GTAACTCTTCGCGTGGTGGGAAACTTTCCGATTTCCATCAAGAGAGTGTATGTGCGCAACG 469
Db 425 AATCTCTTCGCGTACCGGAAATATACCAAAATGTGTTCAAGGCGTTTACGTTTCGAAACG 484
Qy 470 GAGCTAACCACTTCAGAGCCGTCAGAGTCAAGTCAACACTTCTTCGAGGAGAGCGGTATGG 529
Db 485 GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCATTTCTTCGACGCGGAGCGGTATGG 544
Qy 530 TTCACGCGCTCAAAATTCGAACACCGTTTCAGCTAGCTTACGTTTCCCGGTTTACTCAGACTA 589
Db 545 TTCAGCGCGTTCAATTCMAAATGGTGGCTAGTTAGCTTGGCTTTTCACTGAAACAG 604
Qy 590 ACCGTTTGTTCAGAGACGTCATTTGGGTGCGACCGGTTTTTCCCAAGCAATCGGTGAGC 649
Db 605 AGAGGCTTGTTCAGAAAGAAAGCTTTGGGTGCGCCCTTTTCCCTAAAGCAATTTGGTGAAT 664
Qy 650 TTCACGCGCACACCGGTATTGCGCGACTCATGTATTTTACGCGCAGAGCTCGAGCCGGTA 709
Db 665 TACATGCTCACTTGGAAATTTGAAGGCTTATGCTGTTTACGCTCGTGGGCTTTCGGAC 724
Qy 710 TAGTCGACCGGCGACCGAAACCGGTGTAGCTTAACCGCGGTTTGGTCTATTTCATATGGCC 769
Db 725 TTGTTGATCAGCTAAAGGAACTGGTGTTCGAAACCGCGGTTTAGTCTATTTCATTAACC 784
Qy 770 GGTATTGGCTATGTCGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAG 829
Db 785 GATTACTTGTCTATGCTGAAGATGATTGCTTACCCTTAAAGGTAACACCCACCGCG 844
Qy 830 ATTTAAACACCGTTGGTTCGGTTCGATTTTATGAGCAATTTAGAAATCCCAATGATGGCC 889
Db 845 ATCTTAAACAGAGGCTCGATTCGATTCGACCGCCAGCTTAAATCCACATGATAGCTC 904
Qy 890 ACCGAAAGTCGACCGGAAATCCGGTGAATCTTTCGCTTTAGCTACGAGCTGTTTCAA 949
Db 905 ACCAAAGCTCGACCGCAATTTCCGGTGAGCTATTTCCTCTTAGCTACGATGATTTCAGA 964
Qy 950 AGCTTACCTTAAATATCTTCCGATTTCTCAGCGGACGGAATTAATCACCAGCTCGAGA 1009
Db 965 AGCATACCTCAAGTACTTCCAGATTTTCAAAAAATGGGAAAAATCAAATGATGTTGAA 1024
Qy 1010 TTCAGCTTGTATCAGCCAAAGATGATGACGATTTTCGGATTACAGAGAACTTGTGCTG 1069
Db 1025 TTCAGTTGAAGACCCCAACAAATGATGATGATTTTCGCAATTTCTGAGAACTTCGTGCTCA 1084
Qy 1070 TACTGTACCAAGAGTGTGTTTCAAGCTGCGGAGATGATCCCGGTTGGGTCTCCGCTGG 1129
Db 1085 TTCTGTATCAACAGTCTGTTTCAAGATGCTGAATGATCCGTGGAGGTTTCACCGGTGG 1144
Qy 1130 TTTACGACAAGAAACAGGTCGCAAGATTCGGGATTTTACGAAATACGCGGAGATTCAT 1189

```
Db 1145 TTTTACGACAGAACAAAGTTCCCGATTGGTATTCTGGATAGTACGCGAAAGATGGGT 1204
Qy 1190 CGAACATTAGTGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGG 1249
Db 1205 CTCGATTGAATGGGTGAGTACCTGATGTTCTGTTTCCACCTCTGGAATGCTGGG 1264
Qy 1250 AAGAGCCAGAAACAGATGAGTGGTCTGATGAGGTCCTGATGAGTCCACAGACTCAA 1309
Db 1265 AAGAGCAGAAACAGATGAAATCGTTGTAATTGGTTTATGATGATGACACACAGACTCCA 1324
Qy 1310 TTTTACGACGATCTGACGAGATCTCAAGAGTGTCTGCTCTGAATCCGCTGAATCTCA 1369
Db 1325 TTTTCAATGATGATGAGGCTAAGAGTGTGTTTATCCGAATCCGCTCAATTTGA 1384
Qy 1370 AAACCGGTGAATCAACTCGCGCTCGATCATCTCCAAAGAAATCAACAAAGTCAACCTCG 1429
Db 1385 AAACAGGGAATCAACAAAGAAATCCATATCAATCAAAACCCGGATGAACAAAGTGAATTTAG 1444
Qy 1430 AAGCAGGATGGTCAACAGAAACATGCTCGCGCTGAAACCAATTCGCTTACTTGGCTT 1489
Db 1445 AAGCTGAATGGTGAACCCGAAACAAACTCGAAGGAAACAGAGTATGCTTATTTGGCTA 1504
Qy 1490 TAGCCGAGCGTGGCTTAAAGTCTCAGGATTCGCTAAAGTGTATCTCACTACGAGAGAG 1549
Db 1505 TCGCTGAACCATGSCCAAAAGTTCTCGTTTTCGAAAGTAAACCTGTTCAACGGTGAAG 1564
Qy 1550 TTAAGAAACATCTTACCGGATAACCGTTACGAGAGAGAGCTCTGTTTCTCCCGGAG 1609
Db 1565 TTGAGAAATTCATTTAGTGGTGACAAACATATGCTGGGGAACCTCTTTTACCAAGAG 1624
Qy 1610 -----AAGGAGGAGAGAGAGAGGATACATCTCTGTTTCTGTTACGAGAGAGAG 1663
Db 1625 ACCCCACAGACAGAGAGAGAGAGATGTTTATTTAGCTTTCGTTACGATGAGAGAG 1684
Qy 1664 CATGGAATCGAGTTACAGATAGTTAAACCGCTAGCTTAGAGTTGAGGCAACGGTGA 1723
Db 1685 AATGGAATCAGAACTCAAAATTTTAAACGAATGAGTTTGAAGTTGGAGGCAACTGCA 1744
Qy 1724 AACTTCGCTCAAGGTTCCGTTACGAGATTCACGCTACATTCATCGGAGCGAGTATGG 1783
Db 1745 AGCTTCCATCAGAGTTCCITATGATTTTATGAAATTCATGAAACATTCATAAAGCAATGATTTGG 1804
Qy 1784 CGAAGCAGG 1792
Db 1805 CAAATCAGG 1813

RESULT 10
LENEOXANT 2171 bp mRNA linear PLN 22-MAY-1998
DEFINITION Lycopersicon esculentum mRNA for nine-cis-epoxycarotenoid
dioxigenase.
ACCESSION Z97215
VERSION Z97215.1 GI:2769641
KEYWORDS nine-cis-epoxycarotenoid dioxigenase.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1
REFERENCE Burbridge, A., Grieve, T.M., Jackson, A., Thompson, A. and Taylor, I.B.
AUTHORS Structure and expression of a cDNA encoding a putative neoxanthin
TITLE cleavage enzyme (NCE) isolated from a wilt-related tomato
JOURNAL (Lycopersicon esculentum Mill.) library
J. Exp. Bot. 47, 2111-2112 (1997)
2
REFERENCE Burbridge, A.
AUTHORS Direct Submission
TITLE Submitted (01-JUL-1997) Burbridge A., The University of Nottingham,
JOURNAL Physiology and Environmental Science, Sutton Bonington Campus,
Loughborough, Leicestershire, LE12 5RD, UK
```

```
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 2171)
AUTHORS Burbridge, A.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1998) Burbridge A., The University of Nottingham,
Physiology and Environmental Science, Sutton Bonington Campus,
Loughborough, Leicestershire, LE12 5RD, UK
COMMENT On Jan 13, 1998 this sequence version replaced gi:2243153.
FEATURES
source
location/Qualifiers
1..2171
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/db_xref="taxon:4081"
79..1896
/codon_start=1
/product="nine-cis-epoxycarotenoid dioxigenase"
/protein_id="CAE10168.1"
/db_xref="GI:2769642"
/translat="MATTTSHTATNTWIKTLMSPSKKEFGFASNSILKXNHNROSL
NINSSQAIPILHPKQSSNYQTQKNTTSHPKQENNSSSSTSKNVLQKAAVAL
DAVESALTKEHEPLPKTADPRVQI SGNFAPVPENPVCQSLPVTGKI PKCVGVYVR
NGANLPPEPTAGHHFFDGMVHAVQFQNGSASVACRFTETERLVKALGRVFFKVA
IGELHSGGTAEMLFVARGLFGLVHDSKGTGVANAGLVVFNRLAMSDDLPYHVK
VTFGDLKTEGREDFDQGLASTMTIAHPKLDPVSGELFALSVDVIQKPYLYKPKVSKNG
EKSDNVEIPVEDPTMMHDFAITENFVVI PQQVYVFMSEMRGSGSPVVDKXVSRG
KSLYSEIRLNKLTGKSTRKSI IENPDEQVNLGAMVRNKLGRKRTETAYLAIAPNPK
VSGFAKNLFTGEVKEFIYGDNKYGGEPFLPRDPNSKEEDDGYILAFVHDEKEWKE
LQIYNVMSLKEATVKLPSRPVYGFHGTFINANDLANQA"
BASE COUNT 688 a 435 c 431 g 617 t
ORIGIN
Query Match 45.8%; Score 824.6; DB 8; Length 2171;
Best Local Similarity 69.7%; Pred. No. 8.7e-237;
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;
Qy 170 CTCAGCTCTTCATTCCCTAAGCAATCATCAAACTCTCCGCCATTGTTGTTAAGCCCA 229
Db 263 CTTCAAATATCAACACCAAGAAATATACATCAATTTACACCCAAACAGAAACAACA 322
Qy 230 AAGCCAAAGAAATCCAAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGACGCT 289
Db 323 ACTCCTCTCTTCTTCAACTTCCAACTTCCAACTTGTAGTCAGAAAGCAGCAATGGCTT 382
Qy 290 TGGACGCGCGGAGGGTTTCCTTGTACGACGACGAGAGTACACCGCTTCTCTAAAACGG 349
Db 383 TAGATGCTGTAGAAAGTCTTTAACTAAACATGAACCTTTGCGGAAACAG 442
Qy 350 CTGATCTCTAGTGTTCAGATCGCGGAAATTTTGCTCCGGTGAATGAACAGCCCGTCCGCG 409
Db 443 CCGACCCACGAGTCCAGATTTCTGGGAATTTTGCTCCGGTACCGGAAATCCAGTCTGTC 502
Qy 410 GTAATCTTCCGGTGTTCGAAACTTCCGATTCATCAAGAGAGTGTATGTGCGCAACG 469
Db 503 AATCTCTTCCGGTTCACCGGAAATTAACCAATGTGTCAAGGCGGTTTACGTTGAAACG 562
Qy 470 GAGCTAAACCCACTTCCAGAGCGCGGTGACAGGTGCACCACTTCTTCGACGAGACGATGG 529
Db 563 GAGCTAACCTCTCTTTTGAACCAACCGCGGACACCAATTTCTTCGACGCGGACGATGG 622
Qy 530 TTCAGCGCGTCAAAATCGAACGAGTTCAGCTAGTACGCTTCCCGGTTTACTCAGACTA 589
Db 623 TTCAGCGCGTCAAAATCGAAATGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 649
Qy 590 ACCGGTTTCTTCAGGAAACGTCATTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 742
Db 683 AGAGGCTTGTTCAGAAAGAGCTTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 742
Qy 650 TTCAGCGGACACCGGATATTCGCCAGTCTATGCTATTTACGCCAGAGTCGACGCGGTA 709
Db 743 TACATGCTCACCTGGAATTCGAGGCTTATGCTGTTTACGCTCGTGGGTCTTTCGGAC 802
```

QY 710 TAGTCGACCGGCACACGACCGGTGTAGCTAACGCGGTTTGGTCTATTATTCATCGCC 769
Db 803 TTGTTGATCAAGTAAGAACTGGTGTGCAACGCGGTTTGTCTATTATTCATTAACC 862
QY 770 GGTATTGGCTATGTGCGAGGATGATTACCTTACCAAGTTTCAGATCACTCCCAATGGAG 829
Db 863 GATTACTTGTCTGCTGAAGATGATTGCTTACCATGTAAAGTAACACCCACCGCG 922
QY 830 ATTTAAACACGTTGGTTCGATTCGATTTTGAAGGACAAATTAGATCCACATGATTCGCC 889
Db 923 ATCTTAAACACAGAGGTCGATTCGATTTTGAAGGACAAATTAGATCCACATGATTCGCC 982
QY 890 ACCCGAAGTGCACCGGAAATCCGGTGAATCTTCGCTTTTAAGCTACGACCTCGTTTCAA 949
Db 983 ACCCAAAGCTCGACCCAGTTTCGGTGAATCTTCGCTTTTAAGCTACGACCTCGTTTCAA 1042
QY 950 AGCCTTACTAAATACCTTCGATTTCTCAGCGGACGAACTAATATCACCGGACGTCGAGA 1009
Db 1043 AGCCATACCTCAAGTACTTCAGATTTTCAAAAATGGGAAATTCAAATGATTTGAAA 1102
QY 1010 TTCAGCTTGTATCAGCAACAGATGATGACGATTTTCGATTTACAGAGAACTTCGCTGCG 1069
Db 1103 TTCCAGTTGAAGACCCACATGATGATGATTTTCGCAATTTACTGAGAACTTCGCTGCTCA 1162
QY 1070 TACTTGACAGCAAGTCTTTTCAAGCTGCGGAGATGATCGCGTGGTCTCCGCTGG 1129
Db 1163 TTCTCTGATCAACAGTCTGTTTCAAGATGCTGTAATGATCCGTGGAGTTTCACCGGTGG 1222
QY 1130 TTTACGACAAAGAACAGTTCGCAAGATTCGGGATTTTAGCAAAATACGCCCAAGATTCAT 1189
Db 1223 TTTACGACAAAGAACAGTTCGCAAGATTCGGGATTTTAGCAAAATACGCCCAAGATTCAT 1282
QY 1190 CGAATTAAGTGGATGATGCTCCAGATGCTTCTGCTTCATCTCTGGAACGCTTGGG 1249
Db 1283 CTGATTTGAATGGGTGGAATGATGATGATTTCTGTTTCCACCTCGGAATGCTTGGG 1342
QY 1250 RAGAGCCAGAACAGATGAAGTCGCTGATAGGCTCTGATGATGATGATGATGATGATGAT 1309
Db 1343 AAGAGCCAGAACAGATGAAGTCGCTGATAGGCTCTGATGATGATGATGATGATGATGATGAT 1402
QY 1310 TTTTCAACGAGTCTACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTTACTTGGT 1369
Db 1403 TTTTCAATGAATGATGAAGGCTTAAGAGTGTCTGTCTGAAATCCGCTTACTTGGT 1462
QY 1370 AAACCGGTGAATCAACTCGCTCGATCATCTCAAGAGATCAACAGATCAACAGATCAACCTCG 1429
Db 1463 AAACAGGGAATCAACAGAAATCCATAATCGAAACCCGATGAACAGATGAATTTAG 1522
QY 1430 AAGCAGGATGTCAACAGAAATCTCGCTCGGCGTAAACCAAAATTCGCTTACTTGGT 1489
Db 1523 AAGCTGAATGTGAACCAACAACTCGGAAGGAAACAGATGATGCTTATTGGCTA 1582
QY 1490 TAGCCAGCGCTGGCTTAAGTCTCAGATTCGCTTAAGTTCGATCTACTACTGAGAGAG 1549
Db 1583 TCCTGAACCATGGCCAAAAGTTCTGGTTCGAAAGTAAACCTGTCACCGGTGAAG 1642
QY 1550 TTAAGAAACATCTTACCGCGATACCGTTACGAGGAGAGCTCTGTTTCGCCCGGAG 1609
Db 1643 TTGAGAAATCTATTATGTGACACAAATATGTTGGGGAACCTCTTTTITACCAAGAG 1702
QY 1610 -----AAGAGAGAGAGAGAGAGAGATACATCTCTGTTTCGTCACGACGAGAGA 1663
Db 1703 ACCCAACACGAGGAGAGAGAGATGTTTATATTTTACGATGAGAGAGAGAGAGAGAG 1762
QY 1664 CATGGAATTCGAGTTACAGATGTTTAAACCGCTTACGTTAGAGTTGAAGCAACCGTTA 1723
Db 1763 AATGGAATTCAGACTGCAATTTGTTAAGCAATGATTTGAGTTGAGGCAACTGTGA 1822
QY 1724 AACTTCGCTCAAGGTTCCGTACGAGTTTCAAGGATCACTTCAACGAGCCGATGATTTGG 1783
Db 1823 AGCTTCATCAAGAGTTCCTTATGATTTTATGGAATTCATGGAACATTCATAACGCAATGATTTGG 1882
QY 1784 CGAGCAGG 1792

Db 1883 CAATCAGG 1891
RESULT 11
LES439079 19018 bp DNA linear PLN 13-MAR-2003
LOCUS
DEFINITION
Lycopersicon esculentum nced1 gene for 9-cis-epoxycarotenoid
dioxygenase.
ACCESSION
AJ439079 GI:28974076
VERSION
9-cis-epoxycarotenoid dioxygenase; nced1 gene.
KEYWORDS
Lycopersicon esculentum (tomato)
ORGANISM
Lycopersicon esculentum
Bukariyora; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 Thompson, A.J.
Complementation of notabilis, a tomato mutant deficient in the
abscisic acid biosynthetic enzyme 9-cis-epoxycarotenoid dioxygenase
Unpublished
2 (bases 1 to 19018)
Thompson, A.J.
Direct Submission
Submitted (13-MAR-2002) Thompson A.J., Plant Genetics and
Biotechnology, Horticulture Research International, Wellesbourne,
Warwick, CV35 9EF, UNITED KINGDOM
Revised by author [14-MAY-2002]
Related mRNA entry Z97215.
Location/Qualifiers
1. .19018
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="Money-maker"
/db_xref="taxon:4081"
/chromosome="7"
/map="7640"
/clone="NCEd1.1"
/cvt_family="tomato anionic peroxidase inverted repeat
(TAPIR)"
repeat_region
267..567
/tpt_type="INVERTED
2343..3038
/tpt_type="DIRECT
5304..7570
/genes="nced1"
5304..5311
/genes="nced1"
5334..5337
/genes="nced1"
5404..7570
/genes="nced1"
5482..7299
/genes="nced1"
/function="abscisic acid biosynthesis"
/codon_start=1
/product="9-cis-epoxycarotenoid dioxygenase"
/protein_id="CAD30202.1"
/db_xref="GI:28974077"
/translations="MATTTSHTATNTWIKLSMPSSKRGFPASNSISLLKXOHNQSL
NINSLQAPPIHLFPQSNYQTPKNTTISHPKENNNSSSTSKWNLVQKAAHAL
DAYESALTLEHELPKTDPRVQISGNFAPVPENPCQSLFVTKI PKCVQGVYVR
NGANPLFEPTAGHFFDGMVHAGVQKNGSAYACRTFETELVQEKALGRPVFKA
IGLHGHSGTARLMLFYARGLFGLVDHSGTGVANAGLVYFNNRLLMGSDDDLPYHVK
VTGDLKTGTRDFDQGLKSTMIAPHKLDPSVSGELFALSVDVIQKPYLYKPRFSKNG
EKNDVEI PVEDPTMDEFAITENFVVI PQQVVPKMSMIRGGSFVVDKVKSRFG
ILDKYKDGSDKXWVPCFCFLNANEEASTDEIVIGSCMTTTPDIFNECDGL
KSVLSIRLNLTGKSTKRSIIENPDQNLKAGMVRNKLGRKTKRTIAYIALAIPWPK
VSGFKNULFTGEVKEFIYGDNKYGGELFLPRDPNSKEDDGYILAFVHDEKWKSE
LQTVNAMSLEKLEATKLPSPVPGFHTFINANDLANQA"
7570
/gene="nced1"

[illegible]

REFERENCE	1	Burbidge A., Taylor, I.B. and Thompson, A.	
AUTHORS		Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA	
TITLE		Unpublished	
JOURNAL			
REFERENCE	2	(bases 1 to 2164)	
AUTHORS		Direct Submission	
TITLE		Submitted (02-MAR-2000) Burbidge A., Plant Science Division, The	
JOURNAL		University of Nottingham, Sutton Bonington Campus, Loughborough,	
		Leicestershire, LE12 5SD, UNITED KINGDOM	
FEATURES		Location/Qualifiers	
source		1..2164	
		/organism="Solanum tuberosum"	
		/mol_type="mRNA"	
gene		/db_xref="taxon:4113"	
		1..2164	
		/gene="nced1"	
CDS		73..1887	
		/gene="nced1"	
		/function="cleavage of 9-cis-epoxycarotenoids"	
		/codon_start=1	
		/product="putative 9-cis-epoxycarotenoid dioxygenase"	
		/protein_id="CAB76920.1"	
		/db_xref="GI:7209269"	
		/translation="MATTSHATNTWIKPLSMESPSSKEFGFASNSISLLKNOHNRSLNINSLQAPLHPFKQSNITQPTKSTISHPKQNNSSISKNLNVKQAAAMLD	
		AVEGALTQHELEHLPKPTADPRVQISGNFAFNPENVCQSLPVTGKIPKCVQGVYVRN	
		GANLPFETAGRHFDGDMVHAFQFNKGSASACRFTETEFVQEKALGRPPFKAI	
		GELHGSIGALMLFYPARGLFGLIDHSRGCTGVANAGLVYFNRLILAMSEDDLPYHKV	
		TPDGLKTEGDFDGLQKTMIAHPKLDIPDQVFPKSEMIKRGSPVYVDKNKYRFGI	
		KSNDEIVPEEDPTMMDFAITEKFVIFDQVFPKSEMIKRGSPVYVDKNKYRFGI	
		LDKYAKOSDLKWEVPCPCFHLNWEETDEIVVIGSCMTPPDSIINFECDGLK	
		SVLSIRLNLTKGSTRKAIENDEQVNLKAGMNRNKKUGKTKQAYLALAEFPWK	
		SGFAVDLFTGEVEKFTYDKNYGGEPFLPRDPNSKEEDDGYILAFVHDEKEWTSLS	
		QIVNMTLKLEATVKLPRVPYFGHGTFINANDLANQA"	
BASE COUNT	667 a	437 c	433 g 627 t
ORIGIN			
Query Match	45.0%;	Score 809.4;	DB 8; Length 2164;
Best Local Similarity	70.7%;	Pred. No. 3.4e-232;	
Matches 1093;	Conservative	0; Mismatches 446;	Indels 6; Gaps 1;
QY	254	AGATGAATTTGTTCCAGAGAGCGGCGGCGGAGCGGTTGAGCGCGGAGGGTTTCCTTG	313
DB	338	AGTGGAAATTAGTCAGAAAGCAGCAGCAATGGCTTTAGATGCTGTGAAGTGCTTTAA	397
QY	314	TCAGCCACAGAGAGCTACACCGCTTCTTAAACCGCTGATCTAGTGTTCAGATCGCG	373
DB	398	CAAAACACGAACTCGAGCACCCTTTGCCGAAACACAGCCGACCCAGGTCCAGATTTCTG	457
QY	374	GAATTTTCTCGGTGTAATGAACAGCGCTCCGCGTAATCTTCGGGTGTCGGAAC	433
DB	458	GGAAATTCGCTCGGTACCGGAAACCCAGTTTGCCAACTCTTCGGGTACCGGAAAA	517
QY	434	TTCCCGATTCATCAAGAGGTGATGTGCGCAACGAGAGCTAACCCATTACAGAGCGG	493
DB	518	TACCCAAATGTGTTCAAGCGGTTTACGTTTCGAAACGAGGCTAACCCCTTTTCGAACCA	577
QY	494	TCAGAGTCAACCACTTCTTCAGCGGAGAGCGGTATGTTTCAGCGCTCAAAATTCGAACACG	553
DB	578	CCGCGGAGCGCAATTTCTTCAGCGGAGCGGTATGTTTCAGCGCTCAAAATTCGAACATG	637
QY	554	GTTCACTAGCTACGCTTCCCGGTTTACTCAGACTAACCGGTTTGTTCAGGAAAGTCAAT	613
DB	638	GGTCAGCTAGTTACGCTTCCCGGTTTCACTGAAACAGAGAGGTTTGTTCAGGAAAGGCTT	697
QY	614	TGGGTGCGCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGG	673
DB	698	TGGGTGCGCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGGTTTCCCGG	757
QY	674	GACTCATGTATTCTACGCCAGAGGTGAGCGCGGTATAGTCGACCCGCGGACACGGAACCG	733
DB	758	GGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGACTTATTGATCAGAGTAGAGGAACTG	817

QY 1733 CAAGGGTTCGGTACGAGATTTACGGTACATTCATCGGAGCGGATGATTGTCGGAAGCAGG 1792
 Db 1775 CTCGGTTCCTACGGTTTTCATGGAACCTTCATTCAATCCAAAGGATTGAGAAACAAG 1834

RESULT 14

BD017434 1839 bp DNA linear PAT 27-AUG-2002
 Transgenic plant using neozanthine cleaving enzyme gene.
 DEFINITION
 BD017434
 VERSION
 JP 2001258579-A/6.
 KEYWORDS
 Vigna unguiculata (cowpea)
 ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.

REFERENCE

1. (bases 1 to 1839)
 Iuchi, K., Kobayashi, M. and Shinozaki, K.
 Transgenic plant using neozanthine cleaving enzyme gene
 Patent: JP 2001258579-A 6 25-SEP-2001;
 THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH

COMMENT

OS Vigna unguiculata
 PN JP 2001258579-A/6
 PD 25-SEP-2001
 PF 11-JAN-2001 JP 2001003476
 PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC
 C12N15/09, A01H5/00, C12N5/10, C12N5/02, C12N5/00, C12N5/00 CC
 Transgenic plant using neozanthine cleaving enzyme gene PH
 Transgenic plant using neozanthine cleaving enzyme gene PH
 Location/Qualifiers
 (1). (1839).

FEATURES

Source
 1. 1839
 /organism="Vigna unguiculata"
 /mol_type="genomic DNA"
 /db_xref="taxon:3917"
 BASE COUNT 447 a 595 c 442 g 355 t
 ORIGIN

Query Match 44.4%; Score 798.4; DB 6; Length 1839;

Best local Similarity 70.4%; Pred. NO. 6,7e-229;
 Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;

QY 233 CCAAGAAATCCAACTAAACAGATGAATTTGTCAGAGAGCGGCGGAGCGGTTGG 292
 Db 281 CCAACCAACATTACCTCAAAATGGAACCTTCTCCAGAAAGCGCTGCCACGGCTTGG 340
 QY 293 ACGCGCGGAGGTTCTTTGTGAGCCACGAGAGCTACACCGCTTCTTAAACGGCTG 352
 Db 341 ACCTGTCGAAACGGCGCTCTCTCGACGAGCGCAACACCGCTCCCCCAACGGCGG 400
 QY 353 ATCTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAATGAACAGCCCGTCCGGCGTA 412
 Db 401 ACCGAGGTCCTCAATCCCGGAACTTCGCGCGGTGCGGAGCATGCGCGGATCAAG 460
 QY 413 ATCTCCGGTGTGCGAAACCTTCCCGATTCATCAAGAGATGATATGTGCGCAACGGAG 472
 Db 461 GACTCCCGTGTGCGAAACCTTCCCGATTCATGACGCGGTGTACGTGCGCAACGGTG 520
 QY 473 CTAACCACTTCAGAGCGGTGACAGGTCAACACTTCTTCGAGCGGAGCGGTATGTTTC 532
 Db 521 CCAATCCGCTCTAGAGCCTTGTCGCGGACCACTTCTTCGACGCGGACGCGATGCTCC 580
 QY 533 ACGCGGTCAATTCGAAACAGCGTTACGTAGCTACGTTCCCGGTTTACTTCAGACTAAC 592
 Db 581 ACGCGGTGAAGTTCAAGAACGGCGCGCCAGCTACGCTCCCGCTTCAACGAGACGAGC 640
 QY 593 GGTGTTGTTCAGGACGTCAATTGGTTCGACGGGTTTCCCAAGCCATCGGTGAGCTTC 652
 Db 641 GTCTCTCGCAGGAGAAATCTCTAGCGCCCGCGGTGTTCCCGAAGGCCATCGGGAGCTCC 700

QY 653 ACGGCCACACCGGTATTGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCGGTATAG 712
 Db 701 ACGGCCACTCCGGCATCGCGCGGCTCTCTCTTCTACGCGCGGTCTCTTCGGGCTCG 760
 QY 713 TCGACCGGCGACACGGAACCGGTGTAGCTAAACGCCGGTTTGGTCTATTTCATATGGCGGT 772
 Db 761 TTGATGGGTCCAGGSCATGGCGGTGGCGAACGCCGGTCTCGTCTACTTCAACACCAACC 820
 QY 773 TATTGCTATGTCCGAGGATGATTTACCTTACCAAGTTCAGATCATCTCCATGGAGATT 832
 Db 821 TCTTGCCCATGTCCGAAGACGATTTACCTTACCCACGTGAGAAATCACCCCTAAGCGGACT 880
 QY 833 TAAACACCGTTGTCGGTTCGATTTTGTGGAACAATAGATCCACAATGATGGCCACC 892
 Db 881 TAACCAACCGTTGTCGGTTCGATTTTGTGGAACAATAGATCCACAATGATGGCCACC 940
 QY 893 CGAAGTCGACCCGGAATCCGTTGAACCTTCTGCTTTTAAAGTACGACGTCTTCAAAGC 952
 Db 941 CGAACTGGAACCCCGTCGACGGCGACTCCACGGCTCAGCTACGACGTCATTCAAGAC 1000
 QY 953 CTTACCTTAAATACTTCCGATTTCTACCCGAGCGAACTAAATCACCGGACGTCGAGATT 1012
 Db 1001 CTTACCTTAAATACTTCCGATTTCTACCCGAGCGGCTCAAGTCCCCGACGTGGAATCC 1060
 QY 1013 AGCTTGATCAGCCACGATGATCGAATTCGGAATACAGAACTTCGTCGTCTGATC 1072
 Db 1061 CCCTGAAGGAGCCACCATGATGACGATTTCCCATTAACGGAGAATTTGTCGTCTGCTC 1120
 QY 1073 CTGACGAGCAAGTCGTTTCAAGCTGCGGAGATGATCCGCGTGGGTCTCCGCTGTTT 1132
 Db 1121 CCGACCAAGAGGTGTTCTTCAAACTAACGGAGATGATCACCGCGGTCTCCCGTGGTCT 1180
 QY 1133 ACGACAGAAACAAGGTCGCAAGATTCCGGGATTTTAGACAATAACGCCGAGATTTCATCGA 1192
 Db 1181 ACGACAGAAACAAGGTCGCAAGATTCCGGGATTTTAGACAATAACGCCGAGATTTCATCGA 1240
 QY 1193 ACATTAAGTCGATGATGCTCCAGATGCTTCTGCTTCCATCTCTCGAAGCTTGGGAG 1252
 Db 1241 CGATGCGGTGATCGACGCGCGCGATTTGTTCTGCTTCCACCTCTCTGAAACGGTGGAG 1300
 QY 1253 AGCCAGAAACAGATGAAGTCGTGATAGGTCCTGTATGATCTCCACAGACTCAATTT 1312
 Db 1301 AGCCAGAAACAGATGAAGTCGTGATGATGATGATGATGATGATGATGATGATGATGAT 1360
 QY 1313 TCAACGAGTCTGACGAGATCTCAAGAGTCTCTGCTGTAATCCGCTGAACTCAAAA 1372
 Db 1361 TCAACGAGTCTGACGAGATTTGAAGAGCGTGTCTGACAGATAAGGCTGAATCTGAGGA 1420
 QY 1373 CCGGTGATCAACTCGCGCTCCGATCATCTCCAAAGAGATCAACAGTCAACCTCGAAG 1432
 Db 1421 CCGCAAGTCCACTCGCGGCGCCCATATCTCCGAGCGCG---AACAGTGAACCTGGAAG 1477
 QY 1433 CAGGATGTTCAACAGAAACATGCTCGCGCGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492
 Db 1478 CCGCATGGTGAACAGAAACAAAGCTCGGAAGAGAGCCAGTTTCGCGTATCTGCTCTGG 1537
 QY 1493 CCGAGCGTGCCTTAAAGTCTCAGGATTCGTAAGTTGATCTCCTACTCTGGAAGATTA 1552
 Db 1538 CCGAGCGTGCCTTAAAGTCTCAGGATTCGTAAGTTGATCTCCTACTCTGGAAGATTA 1597
 QY 1553 AGAAACATCTTTACGCGGATTAACCGTTACGAGAGAGCTCTGTTTCTCCCGGAGAG 1612
 Db 1598 AGAAGTATCATGTATGGAGAGAGAGAGTTCCGTTGGGAGAGCTCTGTTCTTCTTCTCC 1654
 QY 1613 GAGGAGAGAGAGAGAGAGATACATCTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 1672
 Db 1655 GCCAAAAAGAGAGAGATGGGTATATCTGGCATTCGTGACGAGAGAGAGAGATGGAAAT 1714
 QY 1673 CCGGATTCAGATAGTTAAACCGCGTGTAGCTTTAGAGGTTGAGCAACCGTTAACTCCGT 1732
 Db 1715 CCGAGCTCAGATTGTGATGCCCAAAATTTAAGCTCGAAGCTTCCATCAAACTCCCT 1774
 QY 1733 CAAGGTTCCGTACGGATTTACGGTACATTTCATCGGAGCGGATGATTTCGCGAGAGCAGG 1792

```

Db      1775  CTGCTGTCCTACGGTTTCATGGAACTTTCATTTCCTCAAGGATTGAGAAACAG 1834
|||||
RESULT 15
AB030293
LOCUS
DEFINITION
  AB030293
  Vigna unguiculata CPRD65 mRNA for neoxanthin cleavage enzyme,
  complete cds.
ACCESSION
  AB030293
VERSION
  AB030293.1
KEYWORDS
  neoxanthin cleavage enzyme.
SOURCE
  Vigna unguiculata (cowpea)
ORGANISM
  Vigna unguiculata
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
  Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
  Vigna.
REFERENCE
  1 (sites)
  Iuchi,S., Kobayashi,M., Yamaguchi-Shinozaki,K. and Shinozaki,K.
  A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase
  involved in abscisic acid biosynthesis under water stress in
  drought-tolerant cowpea
  Plant Physiol. 123 (2), 553-562 (2000)
  2031197
JOURNAL
  MEDLINE
PUBMED
  10859185
REFERENCE
  2 (bases 1 to 2432)
  Iuchi,S.
  Direct Submission
  Submitted (16-JUL-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.;
  3-1-1 Kouyadai, Tsukuba 305-0074, Japan
  Location/Qualifiers
  1..2432
  /organism="Vigna unguiculata"
  /mol_type="mRNA"
  /db_xref="taxon:3917"
  1..2432
  /gene="CPRD65"
  126..1964
  /gene="CPRD65"
  /codon_start=1
  /product="neoxanthin cleavage enzyme"
  /protein_id="BAB11932.1"
  /db_xref="GI:9857290"
  /translations="MPSASNTWENATLPSPPKLPSTSSPTNLLPLRKTSSNTIT
  CSLOTLHPKQYPTSTSTSTATTPTPLKTTIATTPPRTNLSLSTNQLPKQW
  NFOKAAATLDJVTALVSHRKHLPKTRADPRVQIAGNFAPVPHRADQSLPVGK
  IPKIDGVYVRNAGNPLYEPVAGHFFDGGMVHAKFVINGAASYACRTETQLRSQE
  KSLGRPVFPKAIIGELHSGIARLLLYFARGLFGLVDSQGMGVNAGLVYNNHLLA
  MSDDLPHYRITPNGDLTTVGRYDFNGQLNSTIAHPKLDVDFGLHALSYDVQKP
  YLKYFRSPGVKSPDVEIPLKEPTWMHDFATENFVVPDQVWFKLTEMITGSPV
  VYDKNTRSFGILHNKAKDANMRWIDAPDCPGFHLNWAWEETBEVVVIGSCMTFA
  DSIPTCEESLSKVLSTIRLNLRKSTREPIISDAQVNLBAGVNRNKLGRKTPA
  YLALAEHPKPVSGFAYDILISGEVKVMYGEKEGEPFLPLNGKEDDGYILAFVHD
  EKWKSELQIVNAQNLKLEASIKLPSRPVIFGHTFIHSKDLRQK"
BASE COUNT      654 a 711 c 533 g 534 t
ORIGIN
  Query Match 44.4%; Score 798.4; DB 8; Length 2432;
  Best Local Similarity 70.4%; Pred. No. 7.2e-229;
  Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;
  233 CCAAGAAATCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCGTTGG 292
  406 CCAACCAACCATTAACCTCAAAATAGAACTTTCTCCAGAAAGCGCGTCCACGCGCTTG 465
  293 ACGCGCGGAGGGTTCTCTGTCAGCCACGAGAGCTACACCGCTTCCTTAAACGCGTG 352
  466 ACCTGGTCGAACGCGCTCGTCTCGCACGCGCAACACCCGCTCCCAAAACGCGGG 525
  353 ATCTAGTGTTCAGATCCCGGAAATTTTGTCCGGTGAATGAACAGCCGCTCCGGCGTA 412
|||||

```

```

Db      526  ACCGAGGGTCCAAATCGCGGGAACCTTCGCGCGGTGCGGAGCAGTCCGCGGATCAAG 585
QY      413  ATCTTCGGTGGTTCGGAATACTTCCCGATTCCATCAAGAGAGTGTATGTGCGCAACGGAG 472
Db      586  GACTCCGGTGGTTCGGAATAATCCCAATGCAATGACGGCGTGTACGTGCGCAACGGTG 645
QY      473  CTAAACCACTTCACGAGCGGTGACAGGTACCACTTTTTCGACGAGGACGGTATGGTTC 532
Db      646  CCAATCCGCTCTACGAGCGCTGTGCGCGGCGCACCACTTCTTCGACGCGGACGGATGGTCC 705
QY      533  ACGCGTCAAAATTCGAACACGGTTTCAGCTAGCTACGCTTTCGCGGTTCCTCAGACTAAAC 592
Db      706  ACGCGTGAAGTTTCAAGACGGCGCGCGCAGCTACGCTGCGCTCCACCGAGACGACG 765
QY      593  GGTTTGTTCAGGAACGTCAAATGGGTGACCGGTTTTTCCCAAGCAATCGGTAGCTTC 652
Db      766  GTCTCTCGCAGGAAATCTCTAGGCGCGCGGTGTTCGGAAGGCGCATCGGGAGCTCC 825
QY      653  ACGSCACACCGGTATTGCGCGCTCATGTATTCTACGCGAGCTGCGAGCGCGGTATAG 712
Db      826  ACGCCCACTCGGCAATCGCGCGCTCTCTCTTCTACGCGCGGTCTCTTCGGGCTCG 885
QY      713  TCGACCCCGCACACGGAACCGGTGTAGCTAACCGCGGTTTGGTCTATTCAATGGCGGT 772
Db      886  TTGATGGGTCCAGGGCATGGCGGTGGCGAAGCGCGGTCTCTCTTCTTCAACAACACC 945
QY      773  TATTGGCTATGTGCGGAGGATTTTACCTTACCAAGTTCCAGATCACTCCCAATGGAGATT 832
Db      946  TCTTGGCCATGTCCGAAGACGATTTTACCTTACCAAGTGAATCAACCTTAAACGGGACT 1005
QY      833  TAAACACCGTTTGGTTCGATTTTGATGGAATTTAGAAATCAACAATGATTGCCCAACC 892
Db      1006  TAAACACCGTTTGGCGGTACGACTTCAACGGGAGCTCAACTCAACAATGATGCCCAACC 1065
QY      893  CGAAGTCGACCCCGGAATCGGTGAATCTTTCGTTTAAAGTCAGAGCTGTTTCAAGC 952
Db      1066  CGAACTGGACCCCGTCGACGCGACCTTCAACGGGCGCTCAAGTCCCGCGGCTCCCAATCC 1125
QY      953  CTTACTTAAATACTTCCGATTCTCACCGGACGGAATTAATCACCGGAGCTCGAGATTC 1012
Db      1126  CTTACTCAAGTACTTCGTTTCTCCCGCGAGCGCTCAAGTCCCGCGGCTCCCGGCTGCT 1185
QY      1013  AGCTTGATCACCACAGATGATGACGATTTTCGGATTTCAGAGAACTTCGTGCTGCTAC 1072
Db      1186  CCCTGAAGAGACCCACCATGATGACCGATTTCGCGCATACGAGGAATTTGTCGCTGCTCC 1245
QY      1073  CTGACCAACAAGTCTGTTTCAAGCTGCGGAGATGATCCGCGGTGGGTCTCCGCTGCTTT 1132
Db      1246  CCGACCAAGTGTGTCTTCAACTAACGGAGATGATCACCGGCGGTCCCGGCTGCTCT 1305
QY      1133  ACGACAAGAACAGGTTCGCAAGATTCGGGATTTTAGACAATATACCGGAGATTCATCGA 1192
Db      1306  ACGACAAGAACAAAACCTCACGGTTTGGGATTCTGCAACAAGAAATGCGAAGACGCGAATG 1365
QY      1193  ACATTAAAGTGAATGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252
Db      1366  CGATTCGGTGTGATCGACGCGCGGATTTGTTCTGCTTCCACCTCTGGAACGCGTGGGAG 1425
QY      1253  ACCCAGAAACAGATGAATCGTCTGTGATAGGGTCTGTATGACTCCACACACTCAATTT 1312
Db      1426  AGCCCGAAACCGAGGAGGTTGTGGTGTATTGGTCTCTGCATACCCCTTCGCGACTCCATTT 1485
QY      1313  TCAACGAGTCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCGCCTCGAATCTCAAAA 1372
Db      1486  TCAACGAATCGAGGAGATTTGAAGAGGTGCTGTACAGATAAGGCTGAACCTTGAGGA 1545
QY      1373  CCGGTGAATCAACTCGCGTCCGATCATCTCCAAAGAAATCAACAAGTCAACCTCGAAG 1432
Db      1546  CCGGCAAGTCCACTCGGCGCCCAATTATCTCCGACGCG--AACAAAGTGAACCTGGGAAG 1602
QY      1433  CAGGGATGTCAACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492
Db      1603  CCGGCATGTGTGAACAGAAACAAAGCTCGGAAGGACCCAGGTTCCGCTATCTGGCTCTGG 1662

```

1493	Qy	CGAGCCGTGGCCTAAAGTCTCAGGATTCGCTAAAGTTCATCTCACTACTCGAGAAAGTTA	1552
1663	Db	CGAGCCCTGGGCCAAAGTCTCGGGCTTTTCGAAAGTTGATTTGCTAGTGGGGAAGTGA	1722
1553	Qy	AGAAACATCTTTACGGCGGATACCGTTACGGAGGAGGCTCTGTTTCTCCCGGAGAA	1612
1723	Db	AGAGTACATGATGAGAAAGAACTTCGGTGGGAGCCTCTGTTTCTCCC---AACG	1779
1613	Qy	GAGGAGGAGAACGACGAGGATACATCCTCTGTTTCGTTACGACGAGAGACATGGAAT	1672
1780	Db	GCCAAAAGAACGATGGGTATATTCTGGCATTCGTGACGACGAGAAAGATGGAAT	1839
1673	Qy	CGAGGTTACAGATAGATTAAACCGCGTTAGCTTAGAGTTGAAGCAACGGTTAACTCCGT	1732
1840	Db	CCGAGCTGCAGATTGTGAATGCCCAAATTTAAAGCTCGAAGTTCCTCAACTCCCT	1899
1733	Qy	CAAGGGTTCGTCACGATTTTCAGGTACATTCATCGAGCCGATGATTTTGGCGAAGCAGG	1792
1900	Db	CTCGTCTCCCTACGGTCTTTTCATGGAACATTTTCATCTCCAAAGATTTGAGGAACAAG	1959

Search completed: November 16, 2003, 17:24:03
Job time : 6764 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 09:37:17 ; Search time 506 Seconds

(without alignments)

9602.744 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atggcttctttcaaggcaac.....tggcgagcaggtcggtgtga 1800

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1800	100.0	1800	22 AAD09396	Arabidopsis thalia
2	824.6	45.8	1818	22 AAD09401	Lycopersicon escul
3	798.4	44.4	1839	22 AAD09399	Vigna unguiculata
4	738.2	41.0	1752	22 AAD09394	Arabidopsis thalia
5	606.6	33.7	1815	22 AAD09400	Zea mays neoxanthi
6	528.2	29.3	1734	22 AAD09398	Arabidopsis thalia
7	229.8	12.8	443	21 AAC56678	Eucalyptus grandis
8	204	11.3	492	22 AAC82706	Rice abscisic acid

9	190.4	10.6	393	25 ABX20484	Human GDP-mannose
10	188	10.4	372	21 AAC56695	Eucalyptus grandis
11	171.8	9.5	325	21 AAC56548	Eucalyptus grandis
12	169.6	9.4	1788	21 AAC42989	Arabidopsis thalia
13	169.6	9.4	1788	22 AAD09395	Arabidopsis thalia
14	169.6	9.4	1788	24 ABZ13639	Arabidopsis thalia
15	169.2	9.4	1950	22 AAF77206	CDNA encoding sunf
16	149.4	8.3	491	21 AAC57157	Pinus radiata tran
17	143.4	8.0	386	21 AAC57162	Pinus radiata tran
18	140.4	7.8	326	21 AAC57167	Pinus radiata tran
19	139.8	7.8	1617	22 AAD09397	Arabidopsis thalia
20	139.8	7.8	1617	22 AAD09402	Arabidopsis thalia
21	139.8	7.8	1777	21 AAC36083	Arabidopsis thalia
22	137.2	7.6	329	21 AAC57165	Pinus radiata tran
23	121.6	6.8	398	21 AAC57145	Pinus radiata tran
24	118	6.6	340	21 AAC56520	Eucalyptus grandis
25	113	6.3	238	21 AAC57169	Pinus radiata tran
26	109.6	6.1	246	21 AAC57168	Pinus radiata tran
27	106.6	5.9	412	21 AAC57164	Pinus radiata tran
28	93	5.2	320	25 ABX20813	Human GDP-mannose
29	89	4.9	190	21 AAC57163	Pinus radiata tran
30	78.6	4.4	303	21 AAC57125	Pinus radiata tran
31	76.8	4.3	308	21 AAC57142	Pinus radiata tran
32	74.8	4.2	395	21 AAC57117	Pinus radiata tran
33	72	4.0	470	24 ABL57537	Apoptosis inhibito
34	67.8	3.8	285	21 AAC56908	Pinus radiata tran
35	65.8	3.7	567	21 AAC56546	Eucalyptus grandis
36	64.6	3.6	457	21 AAC56910	Pinus radiata tran
37	63	3.5	445	21 AAC56948	Pinus radiata tran
38	62.4	3.5	436	21 AAC57124	Pinus radiata tran
39	60.8	3.4	367	24 ABQ85532	Arabidopsis thalia
40	58.2	3.2	304	21 AAC56973	Pinus radiata tran
41	56.6	3.1	419	21 AAC56527	Eucalyptus grandis
42	56.6	3.1	447	21 AAC56480	Eucalyptus grandis
43	56.6	3.1	508	21 AAC56611	Eucalyptus grandis
C	44	3.1	4403765	22 AAI93683	Mycobacterium tube
C	45	3.1	4411529	22 AAI93682	Mycobacterium tube

ALIGNMENTS

RESULT 1

AAD09396

ID AAD09396 standard; cDNA; 1800 BP.

XX AAD09396;

AC AAD09396;

DT 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.
KW Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 1..1800

FT /tag= a

FT /product= "Arabidopsis thaliana AtNCED3 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300219.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Tuchi S, Kobayashi M, Shinozaki K;
PI WPI; 2001-400081/43.
XX DR P-PSDB; AAB04784.
XX PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
XX Claim 3; Page 32-36; 101pp; English.
XX The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzymes play a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An acid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
CC The AtNCE3 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.
XX
XX Sequence 1800 BP; 458 A; 464 C; 439 G; 439 T; 0 other;
Query Match 100.0%; Score 1800; DB 22; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGGTTCTTTTTCACGGCAACGGCTGCGGTTTCTGGAGATGGCTTGGTGGCAATCATACT 60
DB 1 ATGGGTTCTTTTTCACGGCAACGGCTGCGGTTTCTGGAGATGGCTTGGTGGCAATCATACT 60
61 CAGCGCCGATATGCTTCTTCAAGCTCCGACTTGGATTTAGTATTTAGTCTCTTACCTATG 120
DB 61 CAGCGCCGATATGCTTCTTCAAGCTCCGACTTGGATTTAGTATTTAGTCTCTTACCTATG 120
121 GCCAGTCTGTGCACACGTAAGTCAATGTTTTCATCTCGGTTTACACCTCTCCAGCTCTT 180
DB 121 GCCAGTCTGTGCACACGTAAGTCAATGTTTTCATCTCGGTTTACACCTCTCCAGCTCTT 180
181 CATTTCCCTTAGCATCATCAACTCTCCGCCATTTGTTTAAAGCCCAAGCCAAAGAA 240
DB 181 CATTTCCCTTAGCATCATCAACTCTCCGCCATTTGTTTAAAGCCCAAGCCAAAGAA 240
241 TCCAACTAAACAGATGAATTTGTTTCCAGAGAGCGCGGCGAGCGGTTGGACGGCGG 300
DB 241 TCCAACTAAACAGATGAATTTGTTTCCAGAGAGCGCGGCGAGCGGTTGGACGGCGG 300
301 GAGGTTTCTTGTGACGACGAGAGAGTACACCGGTTCTTAAACGGCTGATCTTAGT 360
DB 301 GAGGTTTCTTGTGACGACGAGAGAGTACACCGGTTCTTAAACGGCTGATCTTAGT 360
361 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCGCTCCGCGGCTAATCTTCG 420
DB 361 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCGCTCCGCGGCTAATCTTCG 420
421 GTGGTTCGGAACATTTCCGATTCATCAAGAGTGTATGTGGCCAGCGAGTACACCA 480
DB 421 GTGGTTCGGAACATTTCCGATTCATCAAGAGTGTATGTGGCCAGCGAGTACACCA 480
481 CTTACGAGCGCGGTGACAGGTCAACACTTCTTCGACGAGACGGTATGTTTACGCGGTC 540
DB 481 CTTACGAGCGCGGTGACAGGTCAACACTTCTTCGACGAGACGGTATGTTTACGCGGTC 540
541 AAATTCGAACACGGTTACAGTAGCTACGTTGCGGGTTTACTCAGACTAACCGGTTTGT 600

541 AAATTCGAACACGGTTACAGTAGCTACGTTGCGGGTTTACTCAGACTAACCGGTTTGT 600
601 CAGGAACGTCAATTTGGGTGACCGGTTTCCCAAGACCATCGGTGAGCTTTCACGGCCAC 660
601 CAGGAACGTCAATTTGGGTGACCGGTTTCCCAAGACCATCGGTGAGCTTTCACGGCCAC 660
661 ACCGGTATTCGCGGACTCATGCTATTCTACGCCAGAGCTGCAGCCGGTATAGTCGACCG 720
721 GCACACGAACCGGTTAGCTAAACGCGGTTTGGTCTATTCTTCAATGCCGCTTATTGGCT 780
721 GCACACGAACCGGTTAGCTAAACGCGGTTTGGTCTATTCTTCAATGCCGCTTATTGGCT 780
781 ATGTCGAGAGATGTTTACCTTACCAAGTTAGATCATCTCCCAATGGAGATTTAAAGAC 840
781 ATGTCGAGAGATGTTTACCTTACCAAGTTAGATCATCTCCCAATGGAGATTTAAAGAC 840
841 GTTGGTTCGGTTTGGATTTGATGGACCAATTAGAATCCCAATGATTCGCCACCGCAAGTC 900
841 GTTGGTTCGGTTTGGATTTGATGGACCAATTAGAATCCCAATGATTCGCCACCGCAAGTC 900
901 GACCCGAATTCGGGTGAACTCTTCGGCTTTAAGTAGACGCTGCTTTCAAGGCTTACCTA 960
901 GACCCGAATTCGGGTGAACTCTTCGGCTTTAAGTAGACGCTGCTTTCAAGGCTTACCTA 960
961 AAATACTTCCGATTTCTCACCGGACCGAATCTAAATCACCGGACGTCGAGATTCAGCTT 1020
961 AAATACTTCCGATTTCTCACCGGACCGAATCTAAATCACCGGACGTCGAGATTCAGCTT 1020
1021 CAGCCACGATGATGCAGATTTCCGGATTCAGAGATTCAGAGAACTTCGTCGCTACTGACCA 1080
1021 CAGCCACGATGATGCAGATTTCCGGATTCAGAGATTCAGAGAACTTCGTCGCTACTGACCA 1080
1081 CAAGTCGTTTCAAGCTGCGGAGATGATCCGGGTGGGTCTCCGGTGGTCTCCGGTGGT 1140
1081 CAAGTCGTTTCAAGCTGCGGAGATGATCCGGGTGGGTCTCCGGTGGTCTCCGGTGGT 1140
1141 AACAGGTTCGAGATTCGGGATTTAGACAAATACGCCGAAGATTCATCGAACATTAAG 1200
1141 AACAGGTTCGAGATTCGGGATTTAGACAAATACGCCGAAGATTCATCGAACATTAAG 1200
1201 TGGATTGATCTCCAGATTCCTTCTGCTTCCATCTCTGAAACGCTTGGGAAGAGCCAGAA 1260
1201 TGGATTGATCTCCAGATTCCTTCTGCTTCCATCTCTGAAACGCTTGGGAAGAGCCAGAA 1260
1261 ACAGATGAAGTCGCTGATAGGTCCTGTATGACTCCACGACCTCAATTTTCAACGAG 1320
1261 ACAGATGAAGTCGCTGATAGGTCCTGTATGACTCCACGACCTCAATTTTCAACGAG 1320
1321 TCTGACGAGATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGAA 1380
1321 TCTGACGAGATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGAA 1380
1381 TCRACTCCGCTCGATCTCCAGGAGATTCACAGTCAACCTCCGACGAGGATG 1440
1381 TCRACTCCGCTCGATCTCCAGGAGATTCACAGTCAACCTCCGACGAGGATG 1440
1441 GTCAACGAACATGCTCGGCGGTAACCAATTCGCTTACTTGGCTTTAGCCGCGG 1500
1441 GTCAACGAACATGCTCGGCGGTAACCAATTCGCTTACTTGGCTTTAGCCGCGG 1500
1501 TGGCTTAAGTCTCAGGATTCGATTAAGTTGATCTCACTACTGGAGAGTTTAAAGAACAT 1560
1501 TGGCTTAAGTCTCAGGATTCGATTAAGTTGATCTCACTACTGGAGAGTTTAAAGAACAT 1560
1561 CTTTACGCGATACCGTTACGAGGAGGCTCTGTTTCTCCCGGAGAGAGGAGAG 1620
1561 CTTTACGCGATACCGTTACGAGGAGGCTCTGTTTCTCCCGGAGAGAGGAGAG 1620
1621 GAGACGAGGATACCTCTCTGTTTCAGAGAGAGACATGGAATTCGAGTTA 1680

Db 1621 GAAGACGAAGGATACATCTCTGTTTCCTTCACGACGAGAAGACATGGAATCGGAGTTA 1680
 Qy 1681 CAGATAGTTAAGCGCGTTAGCTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1740
 Db 1681 CAGATAGTTAAGCGCGTTAGCTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1740
 Qy 1741 CCGTACGGATTTACGGTACATTCATTCGAGCCGATGATTTGGCGAAGCAGGTCGTGTGA 1800
 Db 1741 CCGTACGGATTTACGGTACATTCATTCGAGCCGATGATTTGGCGAAGCAGGTCGTGTGA 1800

RESULT 2

AAD09401

ID AAD09401 standard; cDNA; 1818 BP.

XX AC

XX AAD09401;

XX DT 10-SEP-2001 (first entry)

XX DE

Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 cDNA.

XX KW

Tomato; neoxanthin cleavage enzyme; LeNCED1; abscisic acid; ABA;
 stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 plant growth protectant; herbicide; ss.

XX KW

Lycopersicon esculentum.

XX OS

Lycopersicon esculentum.

XX FH

Key Location/Qualifiers

XX FT

1..1818

XX FT

/*tag= a

XX FT

/product= "Lycopersicon esculentum LeNCED1 protein"

XX PN

EP1116794-A2.

XX XX

18-JUL-2001.

XX XX

11-JAN-2001; 2001BP-0300218.

XX PR

13-JAN-2000; 2000JP-0010056.

XX PR

11-JAN-2001; 2001JP-0003476.

XX XX

(RIKE) RIKEN KK.

XX PA

Tuchi S, Kobayashi M, Shinozaki K;

XX PI

WPI; 2001-400081/43.

XX DR

P-PSDB; AAE04789.

XX DR

A DNA encoding a protein with a neoxanthin cleavage activity for
 producing transgenic plants with improved or decreased stress tolerance

XX PT

Claim 3; Page 67-71; 101pp; English.

XX PS

XX CC

The invention relates to neoxanthin cleavage enzymes and their
 corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 plant when expressed in a plant cell. The invention also relates to
 methods for increasing or decreasing stress tolerance in a plant by
 introducing the DNA into the plant, and a transgenic plant into which a
 neoxanthin cleavage enzyme is introduced. The improvement of stress
 tolerance in plants is useful, for example, in plant breeding. Neoxanthin
 cleavage enzyme genes are useful for producing transgenic plants. An arid
 land can be improved by growing transformant weed for several years and
 then removing the weed by specifically lowering stress tolerance in the
 weed by inducing an inducible promoter. The present cDNA sequence encodes
 Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 protein
 related to the invention.

XX CC

XX SQ

Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;
 45.8%; Score 824.6; DB 22; Length 1818;

Query Match

Best Local Similarity 69.6%; Pred. No. 4.8e-257;
 Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;

Qy 170 CTCACGCTCTTCATTTCCCTTAAGCAATCATCAAACTCTCCGCCATTTGTTGTTAAAGCCA 229
 Db 185 CTTCAATTTATCAACACCAAGAAATATACAAATTCACCCAAAACAAGAAACAACA 244
 Qy 230 AAGCCAAAGAAATCCAACTAAACAGATGAATTTGTTCCAGAGAGGGGGGGGAGCGT 289
 Db 245 ACTCTCTTCTTCTTCAACTTCCAAAGTGGAATTTAGTGCAGAAAGCAGCAATGGCTT 304
 Qy 290 TGGACGCGGCGGAGGGTTTCTTGTCTCAGCCACGAGAAGCTACACCGCTTCTTAAACGG 349
 Db 305 TAGATGCTGTAGAAAGTGCTTTAACTAAACATGAAGTGAACACCTTTCCGAAACAG 364
 Qy 350 CTGATCCTAGTGTTCAGATCGCGGAAATTTTGTCTCCCGTGAATGAACAGCCGCTCGGC 409
 Db 365 CCGACCCAGAGTCCAGATTTCTGGGAATTTTGTCTCCGTTACCGGAAATCCAGTCTGTC 424
 Qy 410 GTAATCTTCCGGTGTCCGAAACTTCCCGATTCCATCAAAAGGAGTGTATGCGCAACG 469
 Db 425 AATCTCTTCCGGTACCGGAAATTAACCAATGTGTTCAAGGCTTTACGTTTGAAGC 484
 Qy 470 GAGCTAACCCACTTACGAGCCGCTGACAGGTCAACACTTTCTTCGACGAGACGGTATCG 529
 Db 485 GAGCTAACCCCTCTTTTGAACCAACCGCCGACACCAATTTCTTCGACGCGACGGTATCG 544
 Qy 530 TTCACGCGCTCAAAATTCGAACACGGTTCAGCTAGCTACGCTTCCGCTTACTCAGACTA 589
 Db 545 TTCACGCGCTCAAAATTCGAACACGGTTCAGCTAGCTACGCTTCCGCTTACTCAGACTA 604
 Qy 590 ACCGGTTTGTTCAGGAACGTCATTTGGGTGCGACCGGTTTTTCCCAAGGCAATCGGTGAGC 649
 Db 605 AGAGGCTTGTTCAGGAACGTCATTTGGGTGCGACCGGTTTTTCCCAAGGCAATCGGTGAGT 664
 Qy 650 TTCACGCGCTCAAAATTCGAACACGGTTCAGCTAGCTACGCTTCCGCTTACTCAGACTA 709
 Db 665 TACATGCTCACTCTGGAATTCGAAGCTTATGCTGTTTACGCTCGTGGCTTCTCGGAC 724
 Qy 710 TAGTCGACCGCGCACACGGAACCGGTGTAGCTAAACCGCGTTTGGTCTATTTCAATGGCC 769
 Db 725 TTGTTGATCACAGTAAAGGAACTGGTGTTCGAAACCGCGTTTGGTCTATTTCAATGGCC 784
 Qy 770 GGTATTGGCTATGTCGGAGGATGATTTACCTTACCAAGTTACAGTACCTCCCAATGGAG 829
 Db 785 GATTACTTGTCTGCTGAAGATGATTTGCTTACCATGTAAAGGTAAACCCACCGCGG 844
 Qy 830 ATTAAAAACCGTTGGTTCGGTTTCGATTTGATCGCAATTTAGATCCAAATGATGTCGC 889
 Db 845 ATCTTAAACAGAGGTTGCTGATTCGATTTTCGACGCGCAGCTAAATCCCAATGATGCTC 904
 Qy 890 ACCGAAAGTCGACCCCGAATCCGGTGAATCTTTCGCTTTTAAAGTACGACGCTGTTCAA 949
 Db 905 ACCGAAAGTCGACCCCGAATCCGGTGAATCTTTCGCTTTTAAAGTACGACGCTGTTCAA 964
 Qy 950 AGCTTACCTAAATATCTTCGATTTCTCAGCGACGGAATCAATCACCGAGCTCGAGA 1009
 Db 965 AGCATACCTCAAGTACTTCAAGTTTCAAAAATGGGAAATCAATGATGTTGAA 1024
 Qy 1010 TTCAGCTTGTATCGACCAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1069
 Db 1025 TTCAGCTTGTATCGACCAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1084
 Qy 1070 TACCTGACGAGCAAGTCTGTTTCAAGTTCGCGAGATGATCCGGTGGTCTCCGGTGG 1129
 Db 1085 TTCTGTATCAACAGTCTGTTTCAAGTTCGCGAGATGATCCGGTGGTCTCCGGTGG 1144
 Qy 1130 TTTACGCAAGAAACAAGTTCGCAAGATTTGGGATTTTGAACAATACGCGAAGATTCAT 1189
 Db 1145 TTTACGCAAGAAACAAGTTCGCAAGATTTGGGATTTTGAACAATACGCGAAGATTCAT 1204
 Qy 1190 CGAATTAAGTGTATGATGCTCCAGATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1249

Db 1205 CTGATTGGAATGGGTTGAAGTACCTGATTTCTCTGTTCCACCTCTGGAATGCTTGGG 1264
QY 1250 AAGAGCCAGAAACAGATGAAGTCGTCGTGATAGGTCCTGTATGACTCCACAGACTCAA 1309
Db 1265 AAGAAAGCAGAAACAGATGAATCGTTGTAATGTTGTTATGATGACACACAGACTCCA 1324
QY 1310 TTTTCAACGAGTCTGACGAGATCTCAGAGTGTCTCTGTAATCGGCTGAATCTCA 1369
Db 1325 TTTTCAATGAATGTGATGAAGGCTTAAAGAGTGTATTCGAAATCCGCTCAATTTGA 1384
QY 1370 AAACCGGTGAATCAATCGCGTCCGATCATCTCAACGAGATCAACAAGTCAACCTCG 1429
Db 1385 AAACAGGGAATCAACAAGAAATCCATAATCGAAACCCGGATGAACAAGTGAATTTAG 1444
QY 1430 AAGCAGGATGTCACAGAACATGCTCGCCGTAAACCAATTCGCTTACTTGGCTT 1489
Db 1445 AAGCTGGAATGTTGTAACCGAAACAACTCGGAAGAAACAGATATGCTTATTTGGCTA 1504
QY 1490 TAGCCGAGCGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATPCTCACTACTGGAGAAG 1549
Db 1505 TCGCTGAACCATGGCCAAAGTTTCTGGTTTTCGAAAGTAAACCTGTCACCGGTGAAG 1564
QY 1550 TTAAGAAACATCTTTACGCGGATACCGTTACGGGAGAGCCCTCTGTTCTCCCGGAG 1609
Db 1565 TTGAGAAATTCATTTATGTTGACACAAATATGTTGGGAACCTCTTTTTTTTACCAGAG 1624
QY 1610 -----AAGAGGAGAGAAAGCAGAGATACATCTCTGTTTCGTTACAGCAGAGA 1663
Db 1625 ACCCCACACAGCAAGGAAGACGATGTTATTTAGCTTTCTGTTACAGATGAGAAAG 1684
QY 1664 CATGGAATTCGAGTTACAGATAGTTAAACGCGTTAGCTTAGAGTTTGAAGCAACGTTA 1723
Db 1685 AATGGAATCAGACTCAATTTGTAACCAATGATTTGAAGTTGAGGCAACTGTGA 1744
QY 1724 AACTTCGCTCAGGTTCCGATTCAGGATTTCCAGGTTACATTCATCGAGCCGATGATTGG 1783
Db 1745 AGCTTCCATCAAGAGTTCCTTATGATTTTCATGGAATTCATGAACATTCATGAACGCAATGATTGG 1804
QY 1784 CGAAGCAGG 1792
Db 1805 CAAATCAGG 1813

RESULT 3

AAD09399

ID AAD09399 standard; cDNA; 1839 BP.

XX AAD09399;

XX AC

XX AT

XX CT

XX GG

XX TT

XX 10-SEP-2001 (first entry)

XX Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 cDNA.

XX Cowpea; neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;

XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;

XX plant growth protectant; CowPea Responsive to Dehydration; CPRD65; ss.

XX Vigna unguiculata.

XX Key

XX Location/Qualifiers

XX 1..1839

XX /tag= a

XX /product= "Vigna unguiculata CPRD65 protein"

XX EF1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX

XX

XX

XX

XX

(RIKE) RIKEN KK.

Iuchi S, Kobayashi M, Shinozaki K;

WPI; 2001-400081/43.

P-PSDB; RAB04787.

A DNA encoding a protein with a neoxanthin cleavage activity for producing transgenic plants with improved or decreased stress tolerance

Claim 3; Page 53-56; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present cDNA sequence encodes Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 (CowPea Responsive to Dehydration) protein. CPRD65 gene is isolated from cowpea plant.

Sequence 1839 BP; 447 A; 595 C; 442 G; 355 T; 0 other;

Query Match 44.48; Score 798.4; DB 22; Length 1839;

Best Local Similarity 70.44; Pred. No. 1.6e-248;

Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;

QY 233 CCAAGAATCAACACATAAAGATGAAATTTGTTCCAGAGAGCGCGCGGAGCGTTCG 292

Db 281 CCAACCAACCAATACCTCAAAATGGAATCTTCTCCAGAAAGCGGTGCCCGCTTGG 340

QY 293 ACGCGGCGGAGGTTTCTTGTGAGCCAGAGAGTACACCGCTTCTTAAACGGGCTG 352

Db 341 ACCTGGTGAACCGCGCTCGTCTGCGACGAGCGCAACACCGCTTCCCAAAAGCGCG 400

QY 353 ATCTAGTGTTCAGATCGCGGAAATTTGCTCCGTTGAATGAACAGCCGCTCCGCGTGA 412

Db 401 ACCGAGGGTCCAAATCGCGGAACTTCGCGCGGTGCGGAGCATGCGCGGATCAAG 460

QY 413 ATCTCCGGTGTGCGGAAACTTCCGATTCATCAAGAGTGTATGTGCGCAACGGAG 472

Db 461 GACTCCCGTGTGCGGAAATATCCCAATGCAATTCAGCGGTGACGTGCGCAACGGTG 520

QY 473 CTAAACCACTTCAGACCGCGTGAAGTCAACACTTCTTCGACGAGACGGTATGGTTC 532

Db 521 CCAATCCGCTTACGAGCTGTGCGCGGCAACCACTTCTTCGACGCGGACGGCATGGTCC 580

QY 533 ACGCGTCAATTCGAACAGGTTTACGTAGCTACGTTCCCGGTTTACTTCAGACTAAC 592

Db 581 ACGCGTGAAGTTTCAACAGCGCGCCAGCTACGCTGCCGCTTCCGAGACCGCAGC 640

QY 593 GGTGTGTTTACGAACTGCAATTTGGTGTGACCGGTTTCCCAAGCCATCGGTGAGCTTC 652

Db 641 GTCTCTCGCAGGAAATCTCTAGCGCGCGGTGTTCGGAAGGCAATCGGGGAGCTCC 700

QY 653 ACGGCCACCGGTATTGCGCGACTCATGCTATTCTACGCGAGAGTGTGAGCGCGGTATAG 712

Db 701 ACGGCCACTCCGGCATCGCGGCTCTCTCTTCTACGCGCGGTCTCTTGGGCTCG 760

QY 713 TCGACCGCGCACACGGAACCGGTGTAGCTAACCGCGTTTGGTCTATTCTCAATGGCGGTT 772

Db 761 TTGATGGGTCCCGGCAATGGCGGTGGCGAAGCCCGGTCTCGTCTACTTCAACCAACCC 820

QY 773 TATTGGCTATCTCGGAGGATGATTTACTTTACCAAGTTTCAGATCACTCCCAATGGAGATT 832

DT	10-SEP-2001	(first entry)
XX		
DE	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 cDNA.	
XX		
KW	Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;	
KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
KW	plant growth protectant; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
XX	Key	Location/Qualifiers
XX	CDS	1..1752
FT	/*tag= a	
FT	/product= "Arabidopsis thaliana AtNCED1 protein"	
XX		
PN	EP116794-A2.	
XX		
XX	18-JUL-2001.	
XX		
XX	11-JAN-2001; 2001EP-0300218.	
XX		
XX	13-JAN-2000; 2000JP-0010056.	
PR	11-JAN-2001; 2001JP-0003476.	
XX		
XX	(RIKE) RIKEN KK.	
XX		
PI	Iuchi S, Kobayashi M, Shinozaki K;	
XX		
XX	WPI; 2001-400081/43.	
DR	P-PSDB; AAE04782.	
XX		
XX	A DNA encoding a protein with a neoxanthin cleavage activity for	
PT	producing transgenic plants with improved or decreased stress tolerance	
PT		
XX	Claim 3; Page 18-22; 101pp; English.	
XX		
PS	The invention relates to neoxanthin cleavage enzymes and their	
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key	
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.	
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a	
CC	plant when expressed in a plant cell. The invention also relates to	
CC	methods for increasing or decreasing stress tolerance in a plant by	
CC	introducing the DNA into the plant, and a transgenic plant into which a	
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress	
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin	
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid	
CC	land can be improved by growing transformant weed for several years and	
CC	then removing the weed by specifically lowering stress tolerance in the	
CC	weed by inducing an inducible promoter. The present cDNA sequence encodes	
CC	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 protein.	
CC	The AtNCED1 cDNA is obtained from an Arabidopsis plant-derived cDNA	
CC	library using a cDNA of the CPR65 (Cowpea Responsive to Dehydration)	
CC	gene isolated from cowpea plant as a probe.	
XX		
SQ	Sequence 1752 BP; 484 A; 396 C; 428 G; 444 T; 0 other;	
Query Match		
Best Local Similarity 41.0%; Score 738.2; DB 22; Length 1752;		
Matches 1063; Conservative 0; Mismatches 478; Indels 18; Gaps 2;		
QY	244	AACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGAGCGTTGGACGCGCGAG 303
Db	193	AATCTCTCCGCTAAACATCTTCCAGAAAGCGCGGCGATTCGATCGACGCGCTGAG 252
QY	304	GGTTTCCTTGTCCAGCCAGGAGAGCTACACCGCTTCTCTAAACCGCTGATCCTAGTGT 363
Db	253	CGTGCAATTAATCTCACAGGAGCAAGATTCTCCACTTCCCAAAACCGCTGATCCACGTT 312
QY	364	CAGATCGCGGAAATTTTGTCTCCGTTGAATGAACAGCGCGCTCCGCGGTAATCTCCCGGTG 423
Db	313	CAGATTGCGGGAATTATTCCCGGTACCGGAATCTTCGTCGCGCGGAAACCTCACCGTC 372

821

TCTTGCCATGTCCGAAGACGATTACCTACCACTGAGATCAGCCCTAACCGCGACT

880

833

TAATAACCGTGTGGTTCGATTTTGTAGTGGACAAATTAGAAATCCAAATGATTCGCCACC

892

881

TAACACCGTGTGGCGTTACGACTTCAACGGGAGCTCACTCAACATGATCGCCACC

940

893

CGAAAGTCGACCCCGGAATCCGGTGAATCTTTCGTTTAAAGTACGACGCTGTTCAAAGC

952

941

CGAAACTGGACCCCGCTCGACGCGGACCTCCACGCGCTCAGCTACGACGCTATTGAGAGC

1000

953

CTTACTCTAAATACTTCTCCGATTTCTCACCGGACGGAATTAATCAACGAGCTCGAGATTC

1012

1001

CTTACTCTAAGTACTTCTCGTTTCTCCCGACGGGCTCAAGTCCCGGACGTGGAATCC

1060

1013

AGCTTGATCAGCAACGATGATGACGATTTGCGGATTTACAGAGAACTTCGTCGCTGATC

1072

1061

CCCTCAAGGAGCCCACTGATGACGATTTCCGCATAACGAGAAATTTCTGTCGTCGCTC

1120

1073

CTGACGAGCAAGTCGTTTCAAGCTCCGAGATGATCCGGTGGGTCTCCGGTGGTT

1132

1121

CCGACGAGCAGTGTCTTCAAACTAACGGAGATGATCACCGGCGGTCCCGTGGTCT

1180

1133

ACGACAAGAACAGGTGCGCAAGATTCGGGATTTTACGAAATACGCGGAAGATTCATCGA

1192

1181

ACGACAAGAACAACTCAAGCTTGGATTTCTGCACAAATGCGAAGACGCGAATG

1240

1193

ACATTAAAGTGAATGATGCTCCAGATTTGCTTCTGCTTCCATCTCTGGAACGCTTGGAG

1252

1241

CGATCGGTGGATCGACGCGCGGATTTGTTTCTGCTTCCACTCTCGAAGCGTGGAGG

1300

1253

AGCCAGAAACAGATGAAGTCGTCGTGATAGGTCCTGTATGACTCCACGAGACTCAATTT

1312

1301

AGCCCGAAACGAGGAGTTGTGTGATTTGGTCTGTCATGACCCCTCGGACTCCATTT

1360

1313

TCAACGAGTCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTGATCTCAAA

1372

1361

TCAACGAATCCGAGGAGATTTGAAGAGCGTGTGTCTCAGAGATAAGGCTGAATCGAGA

1420

1373

CCGTGTAATCACTTCGCGCTCCGATCATCTCCAAAGAGATCAACAAGTCAACCTCGAAG

1432

1421

CCGCAAGTCCACTCGCGGCCCATTTATCTCCGACGCCG---AACAGTGAACCTGGAG

1477

1433

CAGGATGTTCAACAGAAACATGCTCGCGCGTAACCAAAATTCGCTTACTTGGCTTAG

1492

1478

CCGCGATGTTGAACAGAAACAAGCTCGGAAGGAGACCCAGTTCGGGTATCTGGCTCTGG

1537

1493

CCGAGCGTGGCTTAAAGTCTCAGATTCGCTAAAGTTGATCTCACTACTGGAAGATTA

1552

1538

CGHAGCCCTGGCCCAAGTCTCGGCTTTGCGAAGTTGATTTGCTGAGTGGGAGTGA

1597

1553

AGAAACATCTTTACGCGGATACCGTTACGGAGAGAGCCTCTGTTTCTCCCGGAGAG

1612

1598

AGAAGTACATGTATGAGAGAGAGAGTTCGGTGGGAGCCTCTGTTTCTTCCC---AAGC

1654

1613

GAGGAGGAGAGACGAGGATACATCTCTGTTTCTGTTTCAAGGAGGAGACATGGAAT

1672

1655

GCCAAAAGAGAGACGATGGGTATATTCTGGCAATCTGTGCACGACGAGAAAGATGGAAT

1714

1673

CGGAGTTACAGATAGTTAAACCGGTTAGCTTGAAGGTTGAAGCAACGGTTAACTTCCT

1732

1715

CCGAGCTCAGATTTGTAATGCCCAAAATTTAAAGCTCGAAGCTTCCATCAAACTCCCT

1774

1733

CAAGGGTTCGTACGATTTTACGAGTACATTCATCGGAGCGCGATGATTTGCGAAGCAGG

1792

1775

CTCGTGTTCCTACGGTTTTCATGGAACTTTTCATTCATTCAGGATTTGAGGAAACAAG

1834

RESULT 4

AAD09394

ID AAD09394 standard; cDNA; 1752 BP.

XX

XX

AC

XX

CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes
 CC Zea mays neoxanthin cleavage enzyme, VP14 protein related to the
 CC invention.

XX SQ Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 other;
 Query Match 33.7%; Score 606.6; DB 22; Length 1815;
 Best Local Similarity 63.5%; Pred. No. 4.3e-186;
 Matches 1003; Conservative 0; Mismatches 549; Indels 27; Gaps 4;

QY 246 CACTAAACAGATGATTTGTCAGAGAGCGCGGCGAGCGTTCGAGCGCGCGGAGG 305
 Db 240 CAAGAGCAGCTCACTTGTTCAGCGCGCGCGGCGCGCTCGACCGTTCGAGGA 299
 QY 306 TTTCCTTGT-----CAGCCACAGAGAGCTACACCCGCTTCCTAAACCGCTGATCTAG 359
 Db 300 AGGTTGTTGGCCACAGCTCTCGAGCGGCCCCACCGGCTGCCAGACGCGCGACCGGC 359
 QY 360 TGTTTCAGATCGCGGAAATTTGTCGCGTGAATGAACAGCGCGTCCGGGTGATCTTCC 419
 Db 360 CGTCAGATCGCGCACTTCGCGCGCGCTCGGGGAGAGCGCGCTGCACGAGCTCCC 419
 QY 420 GGTGTCGGAATCTTCCGATTCCTCAAGAGAGTGTATGTGCGCAACGAGGCTAACCC 479
 Db 420 CGTCTCCGCGCGATCCCGCTTCATCGAGCGGTCTACGCGCAACGCGCGCAACCC 479
 QY 480 ACTTCAGAGCGGTCAGATCACTTTCGAGAGAGCGGTATGTTTCAGCGCT 539
 Db 480 CTGCTTCACCCCGTCGCGCGCGCACTCTTCGAGCGAGCGCATGTGTGACGCGCT 539
 QY 540 ---CAATTCGAAACAGCTTCAGTACGCTTGCCTGCGGTTTACTCAGACTAACCGGTT 596
 Db 540 GCGGATACGAAACGCGCGCGCGAGTCTACGCTTGCCTTCAGGAGCGCGCGCT 599
 QY 597 TGTTCAGGAACTCAATTTGGTTCGACCGGTTTTCGCAAGCGATTCGAGTTCACGG 656
 Db 600 GCGCGAGGCGCGGATCGCGCGCGCTTTCGCAAGCGATTCGAGTTCGACGG 659
 QY 657 CCACACCGTATTCGCGACATCATCTATTCTAGCCAGAGTCGAGCGGTATGATCGA 716
 Db 660 GCACTCCGCGATTCGCGCGCTTCGCGCGCGCGCGCGGTTCGCGCTTCGGA 719
 QY 717 CCGGCGACAGGACCGGTTCAGTTCAGCGCGGTTTGGTCTATTTCATTCAGCGGTTAT 776
 Db 720 CCGCTCGCGCGACCGCGGTTCGCGCGCGCTTCGCTTCTTCAACGCGCGCTGCT 779
 QY 777 GGTATGTCGAGGATGATTTACCTTACCAAGTTCAGATCATCTCCCAATGGAGATTAA 836
 Db 780 CGCATGTCGAGGACGACCTCCCTTACCAAGTTCGCGGTTCGCGCGCGGACCTCGA 839
 QY 837 AACCGTTGTCGTTGATTTGATGGAATTTAGAAATCCCAATGATTCGCCACCGAA 896
 Db 840 GACGTCGCGCGTACGATTCGAGCGGCGAGTCGCGTTCGCGCATGATCGCGCACCCAA 899
 QY 897 AGTCGACCGGAAATCCGTTGAACTTCGTTTAACTAGCTACGAGTCGTTTCAAGCCTTA 956
 Db 900 GCTGACCGCGGACCGCGGAGCTCCACGCGTTCAGCTACGAGCTATCAAGAGCGGTA 959
 QY 957 CTTAAATATCTCCGATTCCTACCGGACGGAATTAATCACCGGAGCTCGAGATTCAGCT 1016
 Db 960 CTTCAAGTATCTTACTTCAGGCGCGAGCGCACCAAGTTCGAGAGTCCCGT 1019
 QY 1017 TGATCGACCAACGATGATGACGATTTGCGATTCAGAGAACTTCGTCGTGATCTGA 1076
 Db 1020 GGAGCGACCGCATGATTCACGATTCGCGCATTCAGAGAACTTCGTCGTGATGCGCGA 1079
 QY 1077 CCAGCAAGCTGTTTCAGGTGCGGAGATGATTCGCGGTGGGTCTCCGCTGTTTACGA 1136
 Db 1080 CCACCGAGGTGTTTCAAGTTCAGGAGATGCTTCGCGCGCGGTTCGCGGTGTCGTA 1139
 QY 1137 CAAGAACAGGTGCGAAGATTCGGGATTTTGAACAAATACGCCGAGATTCATCGAAT 1196

Db 1140 CAAGAGAGAGCTGCGGTTTCGCGTCTCCCAAGCAGCGCGCGTTCGAGAT 1199
 QY 1197 TAAGTGATGATGATCTCAAGATTCCTTCTGCTTCCATCTCTGGAACCTTGGGAGAGCC 1256
 Db 1200 GCGTGGGTGACGTCGCGGACTCTTCTGCTTCCACTGTGGAACTGCGTGGGAGGACGA 1259
 QY 1257 AGAAACAGATCAAGTCGTCGTGATAGGTCTCTGATGACTCCACAGACTCAATTTTCAA 1316
 Db 1260 GCGAGCGGCGAGTGTGTGATGCGCTCTCTGATGACCCCGCGACTCATCTTCAA 1319
 QY 1317 CGAGTCTCAGCAGAAATCTCAAGAGTCTCTGTTGAAATTCGCTTGAATCTCAAAACCG 1376
 Db 1320 CGAGTCCGACGAGCGCTTGGAGAGCGTCTGACCGAGATCCGCTGGACGCGCACGG 1379
 QY 1377 TGAATCAACTCGCTCGATCTCTCAACAGAGATCAACAACTCAACCTTCGAAGCAGG 1436
 Db 1380 CCGGTCCACGCGCGCGCTCTGCTCC---CGCGTCGAGAGAGAACTTGGAGTGGG 1436
 QY 1437 GATGTCACAGAAACATCTCGCGCTTAAACCAAAATTCGCTTACTTTGGCTTTAGCCGA 1496
 Db 1437 CATGTGAACGCAACCTGCTGGCGCGAGAGCGGTACGCTACCTCGCGGTGGCGGA 1496
 QY 1497 GCGTGGCTTAAAGTCTCAGGATTCGTAAGTTGATCTCACTACTTGAGAGAGTTAAGAA 1556
 Db 1497 GCGTGGCGCAAGGAGTCCGGCTTCGCCAAGAGAGACTGTCCAGCGCGAGCTCACCA 1556
 QY 1557 ACATCTTACGCGGATTAACCGTTACGAGAGAGAGCTCTGTTTCTCCCGGAGAGAG- 1615
 Db 1557 GTTCGAGTACGCGAGGCGCGTTCGCGCGGAGCGCTGCTTCTCCATCGACCCGCG 1616
 QY 1616 -----GAGAGAGAGAGAGATACATCTCTACTTGAGAGAGTTAAGAA 1661
 Db 1617 CCGCGCGCACCGCGCGGAGAGAGCGGTACGCTTACCTTCTGCTTCCATCGACGAGCG 1676
 QY 1662 GACATGGAATCGGAGTTACAGATAGTTAAACCGCGTTCAGTTAGAGTTGAAAGCAACG 1721
 Db 1677 CCGCGCACGTCGAGCTACTTGTGTCAATCGCGCGAGATCCGCTTGGAGCGCACG 1736
 QY 1722 TAACTTCGTCAGAGGTTCCGTTACGAGTTTCAAGTTCACGTTACGAGCGAGCTGATTT 1781
 Db 1737 TCAGTTCGCTCCCGGCTCCGCTTCGCGCTTCCAGCGCACCTTTCATCACGCGGCGAG 1796
 QY 1782 GCGAAGCAGGTCGCTGA 1800
 Db 1797 CGAGCGCGAGCGCGCTGA 1815

RESULT 6

AD09398
 ID AD09398 standard; cDNA; 1734 BP.

XX AD09398;

XX 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 cDNA.

XX Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;
 XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 XX plant growth protectant; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT 1..1734

FT /*tag= a /product= "Arabidopsis thaliana AtNCED5 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

XX

PF 11-JAN-2001; 2001EP-0300218.
XX
XX
PR 13-JAN-2000; 2000JP-0010056.
PR 11-JAN-2001; 2001JP-0003476.
XX
XX (RIKE) RIKEN KK.
XX
XX
PI Tuchi S, Kobayashi M, Shinozaki K;
XX WPI; 2001-400081/43.
DR P-PSDB; AAE04786.
XX
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
PT
XX
XX
PS Claim 3; Page 46-49; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.
CC The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPRD65 (CowPea Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.
XX
XX Sequence 1734 BP; 464 A; 390 C; 441 G; 439 T; 0 other;
SQ

Query Match 29.3%; Score 528.2; DB 22; Length 1734;
Best Local Similarity 61.0%; Pred. No. 1.4e-160;
Matches 920; Conservative 0; Mismatches 563; Indels 24; Gaps 3;

QY 306 TTTCTTGTGTCAGCCAGAGAGTACACCGCTTCTCTAAACCGGTGATCTAGTTTCA 365
DB 240 TATCGTTATTCCTATGGAGAGAAATCGCCCGCTTCTTAAACGACCGCCGGGTTCA 299
QY 366 GATCGCCGGAATTTTGTCTCCGTTGAATGAACAGCCCGTCCGCGTAAATCTTCCGGTGT 425
DB 300 ATTATCAGGTAACTTCGCTCCGGTTAATGAATCTCCGGTTCAGAACGGTTTAGAAGTGT 359
QY 426 CGGAARACTTCCGATTCATCAAGAGGTATGTGCGCAACGGAGCTAACCCACTTCA 485
DB 360 TGGTCAGATCTCTTGTCTAAAGAGGATTTACATCCGTAACGGTGCAACCCCTAATGT 419
QY 486 CGAGCCGGTGACAGGTACCACTTCTTCGACGAGAGCGTATGTTTCAGCCGTCATAAT 545
DB 420 TCCGCCGTAGCCGACATCATTTATTTGACGGTGACGGAATGANTCAGCCGTTAGTAT 479
QY 546 CGACACAGGT---TCAGTAGTACGTTGCGCGTTTACTCAGACTAACCGGTTTCTTCA 602
DB 480 CGGTTTGTATACACAGGTAGTTACGTTGCGGTACACTAAACAAACCGGTTTGTTC 539
QY 603 GGAACGTCATTTGGGTGACCGGTTTCCCAAGCCATCGGTGAGCTTCACGCGCCACAC 662
DB 540 AGAAACCGCGCTTGGACGATCGGTTTCCCTAAACCAATCGGAGCTTCACGCGCCATTC 599
QY 663 CGGTATTCGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCCGATAGTACACCGCGC 722
DB 600 CGGTCTAGCTCGACTCGCTCTCTTCAACGGCTCGAGCTGGGATCGGTCTAGTGACCGGAC 659
QY 723 ACACGGAACCGGTGTAGTAAACCGGTTTGGTCTATTTCATGGCCGGTTATTGGCTAT 782
DB 660 ACGTGGCATGGCGGTAGCTAACCGCGGTGTGGTGTCTTTCTTTAACGGCAGGTTATTAGCCAT 719

QY 783 GTCCGAGGATGATTATCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAACCGT 842
DB 720 GTCCGAGGATGATCTTCTTACCAAGTGAAGTTCGACGGTCAAGGAGATCTTTGAGACGAT 779
QY 843 TGGTCGGTTCGATTTTGGATGGCAATAGAAATCCCAATGATTCGCCACCCGAAATCGA 902
DB 780 CGACGGTTCGGATTCGATGACCAAGTTCATCTTTCAGTGATAGCGCATCTTAAGTGG 839
QY 903 CCAGGAAATCCGGTGAATCTCTTTCGCTTAAAGTTCAGACGTCGTTTCAAAGCTTACCTAA 962
DB 840 CGGACCCACAGGAGATCTCCATACACTGAGCTACAAACGTTTGAAGAAACCTCATCTCAG 899
QY 963 ATACTTCGATTCCTCAGCGGACGGAATAAATCAGCGGACGTCGAGATTCAGCTTCATCA 1022
DB 900 GTATCTAAATTCACACGCTCGGGGAAAGACACGTCGCTGGAGATCACGCTCCCTGA 959
QY 1023 GCGAAGATGATGACGATTTTCGGATTTACAGAAATCTTCGTCGTCGTCCTGACGACGA 1082
DB 960 ACCAAGCATGATTCATGATTTCCGATAACCGGAAATTTTGTCTTATACCGGATCAGCA 1019
QY 1083 AGTCGTTTCAAGCTCGCGGAGATGATCCGCGTGGTCTCTCCGGTGGTTCACGACAGAA 1142
DB 1020 AATGGTATTCAAATTCGAAATGATTCGGGCGGTCACCGCTTATCTACGTTAAGA 1079
QY 1143 CAAGTCCGAAATTCGGGATTTTAGACAAATACGCCGAGATTCATCGAACATTAAGTG 1202
DB 1080 AAAAATCGGAGATTTGGAGTTTGTCAAAGCAGGATCTGACCGGTCGGATATAAATTG 1139
QY 1203 GATTGATCTCCAGATTTCTTCCTTCATCTCTGAAACGCTTGGAGAGGACGAGAAC 1262
DB 1140 GGTGATGATCCGATTTGTTTCTGTTTCATCTATGGATGCGTGGAGAGAGACCGA 1199
QY 1263 AGATGAAG-----TCGTCGTGATPAGGTCCTGTATGACTCCACGACTCAATTTT 1313
DB 1200 AGAGGAGAGACCGATTCGTCGTAATCGGTCATGATGAGCCACCGGATCTT 1259
QY 1314 CAACGACTCTGACGAGATCTCAAGAGTCTCTGTAATCCGCTGATCTCAAAAC 1373
DB 1260 TAGTGAATCAGAGAACCAACCGGTTGAAATTAAGTGAGATCCGGTTAAACATCGGTAC 1319
QY 1374 CGGTGAATCAACTCCGCTCCGATCATCTCCACGAAAGATCAACAGTCAACCTCGAAGC 1433
DB 1320 AAAAGATCGAACCGTAAGGTATTCGTAAC-----GGAGTGAATTTAGAAC 1367
QY 1434 AGGATGGTCAACAGAACATGCTCCGCGTAAACCAATTCGCTTACTTGGCTTTAGC 1493
DB 1368 GGGTCACATAAACCGTAGTTACGTTGGCCCGGAAAGCCAGTTCGTTTACATAGCAATAGC 1427
QY 1494 CGAGCCGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGAGAGATTA 1553
DB 1428 CGATCTTGGCCCAATTCAGTGCGATTCGAGAGTAGATATACAAACCGCACCGTTTC 1487
QY 1554 GAAACATCTTTACCGGATTAACCGTTACGAGAGAGACCTCTGTTTCTCCCGGAGAGG 1613
DB 1488 AGAGTTTAATTAACCGACCGGTCGTTGGAGAACCGCTGCTTTGTATCCGAGGAGGA 1547
QY 1614 AGGAGAGAGAACGAAAGGATACATCTCTCTGTTTCAGCGAGAGAGACATGGAATC 1673
DB 1548 AGGAGAGAGACAAAGGTTATGTAAATGGGTTTGTGAGAGACGAGAGAGACGAGTC 1607
QY 1674 GGAGTTACAGATAGTTAAACCGCTTACGCTTAGAGGTTGAAGCAACGTTAACTTCGTC 1733
DB 1608 GGAGTTTGTGGTTCGACGCGACGATATGAAGCAAGTCGCGCGGTCGCGCTTCCCGGA 1667
QY 1734 AAGGTTCCGTTACGATTTTCACGGTACATTCATCCGAGCGCATGATGTTTGGCGAGCAGGT 1793
DB 1668 GAGGATCTTATGGTTTCCATGAAACGTTCTGAGCGAGATCAGTTGAAGAACCAAGT 1727
QY 1794 CGTGGA 1800
DB 1728 TTTCTGA 1734

```

RESULT 7
AAC56678
ID AAC56678 standard; DNA; 443 BP.
XX
AC AAC56678;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #549.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
XX
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
XX
PS Claim 1; Page 490; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 443 BP; 71 A; 177 C; 117 G; 77 T; 1 other;

Query Match 12.8%; Score 229.8; DB 21; Length 443;
Best Local Similarity 69.9%; Pred. No. 7.9e-64;
Matches 309; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 620 GACCGGTTTCCCAAGCCATCGGTGAGCTTCACGGCCACACCGGTATTGCCGACTCA 679
D 1 GCCCGCTTCNCNAGGCCATCGGAGCTCCAGGCCACTCCGGCATCGCGCGCTCA 60
QY 680 TGTATTCTACGCAGAGCTGAGCCGGTATAGTCAGCCGCCACACGGAACCGGTGTAG 739
D 61 TGTCTTCTACGCCCGCAGCCCTTTCGGCTCTGTCGACCACCGGAATGGCATGGCGCTG 120
QY 740 CTAAACCGGTTTGGTCTATTTCATGCGCGGTATTGCTATGTCGAGGATGATTTAC 799
D 121 CGAACCGCGGCTCTGTGTACTTCGAGCGGCACTCTCTCGGATGTCCGAGGACGCTCC 180

```


PN WO200053724-A2.
 XX 14-SEP-2000.
 XX 09-MAR-2000; 2000WO-US06112.
 XX 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI; 2000-579369/54.
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT -
 XX
 PS Claim 1; Page 494; 747pp; English.
 XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 372 BP; 60 A; 148 C; 101 G; 63 T; 0 other;
 Query Match 10.4%; Score 188; DB 21; Length 372;
 Best Local Similarity 69.1%; Pred. No. 2.9e-50;
 Matches 257; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
 QY 639 CATCGTGAGCTTCAGGCGCACACCGGTATGCGCGACTCATGCTATCTTACGCGAGGC 698
 DB 1 CATCGGAGAGCTCCACGCGCACTCGGCGATCGCGGCTCATGCTCTTCTACGCGCGAG 60
 QY 699 TGCAGCGCGGTATAGTCGACCGCGCACACGCGGTAGCTAACGCGGTTTGGTCTA 758
 DB 61 CCTCTTCGGCCTCGTGACCCACCGGAATGGCATGGCGTTCGGAGCGCGGCTCGTGA 120
 QY 759 TTTCAATGCGCGGTATGCGGTATGCGGAGATGATTTACCTTACCAAGTTCAGATCAC 818
 DB 121 CTTGACGCGCCACTCCTCGCGATGTCGAGGACGACGACCTCCCTACCACTGCGGTCAC 180
 QY 819 TCCCAATGGAGATTTAAAGACCGTTGGTTCGGTTCGATTTGATGGACAATTAGAAATCCAC 878
 DB 181 GGGCTCGCGGACCTCGAGACCGTTCGGCGGTACGACTTCGCGCGGAGCTCGACTTCC 240
 QY 879 AATGATGCCCCCGGAAAGTCAGACCGCGAATCCGGTGAATCTTTCGTTTAACTACGA 938
 DB 241 GATGATCGCCACCGCGAAGATCGACCGGCTTCGCGGAGATGTTTCGCGCTCAGCTACGA 300
 QY 939 CGTCGTTTCAAGCGCTTACCTAAATACTTCGATTTCTCAGCGGAGAACTAAATCAC 998
 DB 301 CGTCGTCGGAAGCGTACCTCAAGTACTTCGATTTCTCAGGAGCGCGGAGAGTCCCC 360
 QY 999 GGACGTCGAGAT 1010
 DB 361 CGACGTCGAGAT 372

RESULT 11
 AAC56548

ID AAC56548 standard; DNA; 325 BP.
 XX AC AAC56548;
 XX 25-JAN-2001 (first entry)
 DT DE Eucalyptus grandis transcription factor DNA sequence #419.
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX Eucalyptus grandis.
 OS WO200053724-A2.
 XX PN 14-SEP-2000.
 XX PD 09-MAR-2000; 2000WO-US06112.
 XX PF 11-MAR-1999; 99US-0266513.
 XX PR 18-AUG-1999; 99US-0149485.
 XX PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI; 2000-579369/54.
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT -
 XX
 PS Claim 1; Page 461; 747pp; English.
 XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 325 BP; 45 A; 128 C; 100 G; 52 T; 0 other;
 Query Match 9.5%; Score 171.8; DB 21; Length 325;
 Best Local Similarity 71.2%; Pred. No. 5e-45;
 Matches 227; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 QY 451 GGAGTGATGTGCGCAACCGAGCTAACCCATTCACGAGCGCGGTGACAGGTACACACTTC 510
 DB 3 GGGGTGTAGCTCGGAAACGCGCCACACCGCTCCACGAGCGGTGCGCGGACCACTTG 62
 QY 511 TTCGCGGAGACGGTATGTTTTCACGCGGTCAAATTCGAACACGGTTCAGCTAGCTACGCT 570
 DB 63 TTCGCGGAGACGGTATGTTTTCACGCGGTCGCGGTTCGCGGCGGTTCAGTGTAGCTAC 122
 QY 571 TGCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCATTTGGTTCGACCGGTTTC 630
 DB 123 TGCGGTTTACCGAGACGCAACCGCTGATCCAGGAACGGGCTCGCGCGGCTTCCTTC 182
 QY 631 CCCAAGCCATCGGTGAGCTTACGCGCACACGGTATTCGCGGACTGATGTTTCTAC 690
 DB 183 CCCAAGCCATCGGTGAGCTTACGCGCACACGGTATTCGCGGCTGATGTTTCTTC 242

QY 691 GCCAGAGTCGAGCGGTATAGTCACCCCGGCACACGGAACCGGTGTAGCTAACGCGGT 750
Db 243 GCCCGCGCCCTCTTCGGCTTCGACACCGTAATGCGATGGCGTCCGGAACGGCGC 302
QY 751 TTGGTCATTTCATATGGCC 769
Db 303 CTCGTGACTTCGACGCC 321

RESULT 12

AAC42989
ID AAC42989 standard; DNA; 1788 BP.

XX AC AAC42989;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37611.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123160.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136382.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142377.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.

PR	10-AUG-1999;	99US-01481171.	Db	393	CGGCACTCTTCCACTGTCACTTAACGGCGCTTACATCGTAAACGGTCCAAATCCACAGTT	452
PR	11-AUG-1999;	99US-01483119.	Qy	486	CGAGCCGGTGACAGGTCAACCACTTCTTCGACGGAGACGGTATGTTTCACGGCGTCAAAAT	545
PR	12-AUG-1999;	99US-01483411.	Db	453	TCTCCCTCGTGGTCTTACCATCTCTTCGACGGGACGGTATGCTTTCACGCCATAAAAT	512
PR	13-AUG-1999;	99US-0148684.	Qy	546	CGAACACGGTTCAGTCTAGTCTGCGGCTTACTCAGACTAACCGGTTGTTGTCAGA	605
PR	16-AUG-1999;	99US-01491175.	Db	513	CCACAACGGTAAGCCACTCTCTGTAGCAGATACGTCAAGACTTATAAATCAACGTCGA	572
PR	17-AUG-1999;	99US-0149426.	Qy	606	ACGTCAATTTGGGTGACACGGGTTTCCCAAGGCCATCGGTGAGCTTTCACGGSCACACCG-	664
PR	20-AUG-1999;	99US-0149723.	Db	573	GAACAAACCGGAGCTCGGTTATGCTTAACTGTTTCCGGATTCAACGGTGTAAACGC	632
PR	23-AUG-1999;	99US-0149902.	Qy	665	--GTATTGCCCACTCATGCTATTCTACGCCAGACTGACGCCGTATAGTCGACCCGCG	722
PR	25-AUG-1999;	99US-0150566.	Db	633	GTCACTAGTCTCGTAGGCTTTTAAACGGCAGCTAGGCTTTTAAACGGCAGTATAATCCCGG	692
PR	26-AUG-1999;	99US-0150884.	Qy	723	ACACGGAACCGGTAGCTAACCGGTTTGGTCTATTTCATATGCGCGGTTATTGGCTAT	782
PR	27-AUG-1999;	99US-0151066.	Db	693	TAACGGCATGTTTGTAGCTAATACAGCTAGCTTCTTCAGTAACCGCTCTCTTGTCTT	752
PR	27-AUG-1999;	99US-0151066.	Qy	783	GTCCGAGGATGATTACCTTACAAAGTTCAGATCATCTCCCAATGAGATTTAAACCGT	842
PR	27-AUG-1999;	99US-0151066.	Db	753	AGTGAATCTGATTTACCTTACCGGTCGATTAACCGAATCAGGAGATATTGAAACCGAT	812
PR	27-AUG-1999;	99US-0151066.	Qy	843	TGGTGGTTCGATTTTGTAGGACAAATAGATCCCAATGATTGCCACCCGAAAGTCGA	902
PR	27-AUG-1999;	99US-0151066.	Db	813	CGACCGGTACGATTTTCGACGGGAATTAGCGATGATGATGACAGCTCATCTTAAACCGA	872
PR	27-AUG-1999;	99US-0151066.	Qy	903	CCCGGAATCCGGTGAATCTTCCGCTTAAAGTTCAGACGTCGTTTCAAAGCTTACCTAAA	962
PR	27-AUG-1999;	99US-0151066.	Db	873	TCCAAATAACCGGAGAACTTTCGCTTTCGGTACGGTCCGGTTCCA---CCGTTTTTAA	929
PR	27-AUG-1999;	99US-0151066.	Qy	963	ATACTTCCGATCTCACCGGACGGAACTAAATCACCAGCGTCGAGA---TTTCAGTTGA	1019
PR	27-AUG-1999;	99US-0151066.	Db	930	ATATTTCGGTTGATTCGCGCGGGAATAAACAAGAGACGTTCCGATTTCTCGATGAC	989
PR	27-AUG-1999;	99US-0151066.	Qy	1020	TCAGCCAAAGATGATGACGATTTCCGATTCAGAGAACTTCGTCGTCTACCTGACCA	1079
PR	27-AUG-1999;	99US-0151066.	Db	990	GTCTCCGTCTTCTCCATGATCTTCGATACGAAACGTCACCGGATTTTCGACAGAT	1049
PR	27-AUG-1999;	99US-0151066.	Qy	1080	GAAAGTCG-----TTTTCAAGTCGCGGAGATGATCCGCGTGGCTTCCCGTGT	1130
PR	27-AUG-1999;	99US-0151066.	Db	1050	TCAGCTTGGCATGAGGATGAACATGTTGGATTGTTGTTCTCGAAGTGGTTCCTCCGTTGG	1109
PR	27-AUG-1999;	99US-0151066.	Qy	1131	TTACGCAAGAACAGGTCCGAAGATTCGGGATTTTAGACAAATACGCCAGAGATTCATC	1190
PR	27-AUG-1999;	99US-0151066.	Db	1110	TACTGATACGGAATAACTCCAGGCTTGAGTGATTCCTAAGTACGCCGAGATGAGTC	1169
PR	27-AUG-1999;	99US-0151066.	Qy	1191	GAACATTAGTGGATGATGCTCCAGATGCTTCTGCTTCCATCTCTGGAACGCTTGGGA	1250
PR	27-AUG-1999;	99US-0151066.	Db	1170	GGAGATGAATGGTTCGAAATTCCTGGATTCATATCATCTACGCTATTAAATGCTTGGGA	1229
PR	27-AUG-1999;	99US-0151066.	Qy	1251	AGAGCCAGAAACAGATGAAGTCTGCTGATAGGCTCTGATGACCTCCACAGACTCAAT	1310
PR	27-AUG-1999;	99US-0151066.	Db	1230	TGAAGATGATGGAACAGCGTCTGTTTGTATGACCCGAAATATTATGCTGATTAACATAC	1289
PR	27-AUG-1999;	99US-0151066.	Qy	1311	TTTCAACGAGTCTGACGAGAACTCAGAGTGTCTGCTGTGAAATCCGCTGATCTCAA	1370
PR	27-AUG-1999;	99US-0151066.	Db	1290	TTTAGAGAGGATGATCTGGTTTC---ATGCTTTGGTGGAGAGGTGAAGATCGATCTCGT	1346
PR	27-AUG-1999;	99US-0151066.	Qy	1371	AACCGTGAATCAACTCCGCTCCGATCATCTCCACGAAGATCAACAACTCAACCTCGA	1430
PR	27-AUG-1999;	99US-0151066.	Db	1347	CACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATCTCGA	1391
PR	27-AUG-1999;	99US-0151066.	Qy	1431	AGCAGGATGTTCAACAGAAACATGCTCGGCCGCTAAACAAATTCGCTTACTTGGCTTT	1490
PR	27-AUG-1999;	99US-0151066.	Db	1392	TTTCGCTGTGATTAATCCGCGCTTCTCCGGAAGATGATGACAGTACGTTTACCGCGGAT	1451
PR	27-AUG-1999;	99US-0151066.	Qy	1491	AGCCGACCGTGGCTTAAGTTCAGATTCGCTTAAAGTTGATCTCAGTCTACTCTGAGAGT	1550
PR	27-AUG-1999;	99US-0151066.	Db	1452	TGGAGATCCGATGCCGAAGATCTCCCGTGTGGTGAAGCTTGTGTCTTAAGAGATCG	1511

Query Match 9.4%; Score 169.6; DB 21; Length 1788;
Best Local Similarity 50.2%; Pred. No. 7.1e-44;
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;
426 CGGAAACTTCCCGATTCCATCAAGAGGTGATGTGCGCAACGGAGCTTAACCCACTTCA 485

QY 1551 TAAGAAACATC-----TTTACGGGATAAACCGTTACCGAGGAGCCCTCT 1595
Db 1512 GATGATGTACGGTGGCCGCTAGATGTACGGTTTACGGCGGAGAACCGTT 1571
QY 1596 GTTCTCCCGGAGAGGAGGA-----GAGGAACGAGAGATACATCTCTG 1643
Db 1572 TTTCTAGTAGGATCTCTGTTAATCCGAGCGGAGGAGGATGATGTTATGTTGAG 1631
QY 1644 TTTCTGTTACGACGAGAGACATGAAATCGAGTTACAGATAGTTAAACGCGCTAGCTT 1703
Db 1632 GTATGTTACGATGAGTACGTTGAGATCGAAGTTTCTGTTGATGGACGCTAAATCGCC 1691
QY 1704 AGAGGTTGAA-----GCAACGGTTAACTTCGTCAGAGGTTCCGTACGAGTTTCACGG 1757
Db 1692 GGAGCTTGAATTCGTGCGCGCGTGAAGTTGCCGGAAGGGTTCGTTACGAGTTCCATGG 1751
QY 1758 TACATTTCATCGGAGCGGATGATTGGCGAAGC 1789
Db 1752 GTTATTTGTCAGGAAGTGACCTTAATPAGC 1783

RESULT 13

AAD09395
ID AAD09395 standard; cDNA; 1788 BP.
XX
AC AAD09395;
XX
DT 10-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 cDNA.
XX
KW Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.
XX
OS Arabidopsis thaliana.
XX
PH Key Location/Qualifiers
FT CDS 1..1788
FT FT /tag= a
FT FT /product= "Arabidopsis thaliana AtNCED2 protein"
XX
FN EF116794-A2.
XX
XX 18-JUL-2001.
XX
XX 11-JAN-2001; 2001EP-0300218.
XX
XX 13-JAN-2000; 2000JP-0010056.
XX
XX 11-JAN-2001; 2001JP-0003476.
XX
XX (RIKE) RIKEN KK.
XX
XX Iuchi S, Kobayashi M, Shinozaki K;
XX
XX WPI; 2001-400081/43.
XX
XX P-PSDB; AAE04783.

A DNA encoding a protein with a neoxanthin cleavage activity for producing transgenic plants with improved or decreased stress tolerance

Example 10; Page 25-29; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress

CC tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transgenic plant for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present cDNA sequence encodes Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein. The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration) gene isolated from cowpea plant as a probe.

XX Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Query Match 9.4%; Score 169.6; DB 22; Length 1788;
Best Local Similarity 50.2%; Pred. No. 7.1e-44;
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;

QY 426 CGAAATCTCCGATCCATCAAGAGTGTATGCGCAACGGAGCTAACCCACTTCA 485
Db 393 CGGCACTCTTCCACTGTCACTTAACCGCGCTTACATCCGTACGGTCCAAATCCACAGTT 452
QY 486 CGAGCGGTGACAGGTCAACACTTCTTCGACGGAGACGGTATGTTTCACGCGTCAAAAT 545
Db 453 TCTCCCTCGTGGTCCCTTACCATCTCTTCGAGCGGACGGTATGCTTCACGCCATAAAAT 512
QY 546 CGAACACGGTTACGTAGTACGCTTCCCGGTTTACTCAGACTAACCGGTTTGTTCAGGA 605
Db 513 CCACACGGTAAAGCCACTCTCTGTAGCAGATACGTCACAGACTTATAATACACGTCGA 572
QY 606 ACGTCAATTGGGTGACCGCGTTCCTCCCAAGCCATCGGTGAGCTTCACGGCCACACCG - 664
Db 573 GAAACAAACCGGAGCTCGGTTATGCTAACGGTGTTCGGATTCAACGGTTAAACGGC 632
QY 665 --GTATTCGCCGACTCATGCTATTCTACGCCAGAGTGCAGCGGTATAGTCGACCGCGC 722
Db 633 GTCAGTAGCTCGTGGAGCTTTAAACGGCAGCTAGGTTTAAACCGGACAGTATTAATCCGGT 692
QY 723 ACACGGAACCGGTGTAGCTAACCGCGTGTGCTATTTCATATGGCGGTTATTTGGCTAT 782
Db 693 TAACGGCTATGTTTGTAGTAACTACAGTCTAGCTTCTTCCAGTAACCGTCTCTTTGCTTT 752
QY 783 GTCGAGGATGATTTACCTTACCAAGTTTACAGTACATCCCAATGGAGATTTAAACCGT 842
Db 753 AGGTGAATCTGATTTACCTTACCGCTACCGCTCCGATTAAACCGAATCAGGAGATTTGAAACGAT 812
QY 843 TGGTCGGTTGATTTGATGACAAATAGAAATCCAAATGATTCGCCACCGCAAGTCGA 902
Db 813 CGGACGGTACGATTTCCGCGGAAATTTAGGATGAGTATGACAGCTCATCTTAAACCGA 872
QY 903 CCCGGAATCCGGTGAATCTTTCGCTTTAAGCTACGAGCTGCTTTTCAAAGCTTACCTAAA 962
Db 873 TCCAATAACCGGAGAAACTTTCGCTTTCCGGTACGGTCCGGTTCCA---CCGTTTTTAAAC 929
QY 963 ATACTTCGGATTCACCGGACGGAACCTAAATCACCAGGAGTTCGAGA---TTCAGCTTGA 1019
Db 930 ATATTTCCGGTTTGAATTCGCCCGGGAATAACAAGAGAGCTTCCGATATTTCTCGATGAC 989
QY 1020 TCAGCCAAACGATGATGACACGATTTTCGCGATTTACAGAGAACTTCGTCGTCGCTGACCA 1079
Db 990 GTCTCCGTCGTTTCTCCATGACTTTCGGGATTCACGAAACGTCACCGGATTTTCGAGAGAT 1049
QY 1080 GCAAGTCG-----TTTTCAAGCTCCCGAGATGATCCGCGTGGTCTCCGGTGGT 1130
Db 1050 TCAGCTTGGCATGAGGATGAACATGTTGGATTGTTCTCGAAGGTGTTCTCCGGTTGG 1109
QY 1131 TTACACCAAGAAACGAGTTCGCAAGATTCGGGATTTTAGACAAATACCGCGAAGATTCATC 1190
Db 1110 TACTGATTAACGGAAACCTCCAAAGCTTGGAGTGATTCCTAAGTACGCGGAGATGAGTC 1169
QY 1191 GAACATTAAGTGGATGATGCTCCAGATGCTTCTGTTCCATCTCTGGAACGCTTGGGA 1250
Db 1170 GGAGATGAATGGTTTCCGAAGTTCCTGGATTCAATATCATTCACGCTATTATGCTTGGGA 1229
QY 1251 AGAGCCAGAAACAGATGAAGTTCGTCGATAGGGTCTCTGTATGACTCCACAGACTCAAT 1310

Db	1230	TCAGATGATGAAACAGCGTCTTTGATTGACCGAATATTATGTCGATTGAACATAC	1289
Qy	1311	TTTCAACGAGTCTGACGAGATCTCAAGAGTCTCTGTCGTAATCCCGCTGATCTCAA	1370
Db	1290	TTTAGAGAGGATGGATCTGTGTTTC---ATGCTTTGGTGAGAGAGTGGAAGATCGATCTCGT	1346
Qy	1371	AACCGGTGAATCAACTCCCGTCCGATCATCTCCAAACGAAGATCAACAAAGTCAACCTCGA	1430
Db	1347	CACCGGATGTGAGACGTCATCCGATCTCAGCGAGG-----AATCTCGA	1391
Qy	1431	ACGAGGATGGTCAACAGAAACATCTCGCGCGTAAACCCAAATTCGCTTACTTGGCTTT	1490
Db	1392	TTTCGCTGTGATTAATCCGCGCTTTCTCGGGAGATGTAGCAGGTACGTTTACGCGCGAT	1451
Qy	1491	AGCCGAGCGTGGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAAGAT	1550
Db	1452	TGAGATCCGATGCCGAAGATCTCCGGTGTGTGAAGCTTGATGTCTTAAAGGAGATCG	1511
Qy	1551	TTAGAAACATC-----TTTACGGGATTAACCGTTACCGGAGAGAGCCTCT	1595
Db	1512	GGATGATTGTACGGTGGCCGCTAGAAATGTACGGTTACGGTTGTACGGCGAGAACCGTT	1571
Qy	1596	GTTCCTCCCGGAGAGAGGAGGA-----GAGGAAGACGAGGAGATACATCTCTG	1643
Db	1572	TTTCGTAGCTAGGATCCTGGTAAATCCGGAGCGGAGGAGATGATGGTTATGTGTGAC	1631
Qy	1644	TTTCGTTCACGAGAGACATGGAATCCGAGTTACAGATAGTAACGCGCTTAGCTT	1703
Db	1632	GTATGTTACGATGAAGTACGATGGAATCGAAGTTTCTGTGATGGACGCTAAATCGCC	1691
Qy	1704	AGAGGTTGAA-----GCAACGTTTAAACTTCGCTCAAGGTTCCGTTACGGATTTCA	1757
Db	1692	GGAGCTTGAATCGTCCGCGCGTGAGTTGCGCGAAGGTTCCGTACGATTCATCGG	1751
Qy	1758	TACATTCATCGGAGCGGATGATTGGCGAAGC	1789
Db	1752	GTTATTGTCAAGGAAAGTGAACCTTAATAAGC	1783
RESULT 14			
Id	ABZ13639	standard; DNA; 1788 BP.	
XX	AC	ABZ13639;	
XX	DT	21-JAN-2003 (first entry)	
XX	DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 1444.	
XX	XW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.	
XX	OS	Arabidopsis thaliana.	
XX	PN	W020021655-A2.	
XX	PD	28-FEB-2002.	
XX	XX	24-AUG-2001; 2001WO-US26685.	
XX	PR	24-JUN-2000; 2000US-227866P.	
XX	PR	26-JAN-2001; 2001US-264647P.	
XX	PR	22-JUN-2001; 2001US-300111P.	
XX	PA	(SCRI) SCRIPPS RES INST.	
XX	PA	(SYGN) SYNGENTA PARTICIPATIONS AG.	
XX	PI	Harper JF, Krebs J, Wang X, Zhu T;	
XX	XX	WPI; 2002-304127/34.	
XX	DR	Identifying a stress condition to which a plant cell has been exposed	
XX	PT	and producing plants with increased tolerance to these abiotic stresses	

PT	XX	Claim 144; SEQ ID NO 1444; 577pp + Sequence Listing; English.	
PS	XX	The invention relates to identifying a stress condition to which a plant	
XX	CC	cell has been exposed, comprising:	
CC	CC	(a) contacting nucleic acid representative of expressed polynucleotides	
CC	CC	in the plant cell with an array of probes representative of the plant	
CC	CC	cell genome; and	
CC	CC	(b) detecting a profile of expressed polynucleotides in the plant cell	
CC	CC	characteristic of a stress response. The method is useful in the	
CC	CC	production of transgenic plants, cells and seeds and in producing plants	
CC	CC	with increased tolerance to abiotic stress. The present sequence is that	
CC	CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used	
CC	CC	in methods of the invention.	
CC	CC	Note: The sequence data for this patent is not represented in the printed	
CC	CC	specification but is based on sequence information supplied to Derwent by	
CC	CC	the European Patent Office.	
XX	XX	Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;	
Qy	393	CGGAAATCTCCCGATTCCATCAAGAGAGTGTATGTCGCAACGAGCTAAACCCACTTCA	485
Db	393	CGGCACCTCTTCCACTGTCACTTAACGGCGCTTACATCCGTAAACGTCCTCAATCCACAGTT	452
Qy	486	CGAGCGGTGACAGGTGACCACTTCTTCGAGGAGAGCGGTATGTTTCACGCGCTCAAAAT	545
Db	453	TCTCCCTCGTGGTCTTACCCTCTCTTCGACGCGAGCGGTATGTTTCACGCGCTCAAAAT	512
Qy	546	CGAACACGGTTCAGTACGTACGTTGCGCGTTTACTCAGACTAAACCGGTTTCTCAGGA	605
Db	513	CCACACGGTAAGCCACTCTCTGTAGCAGATACGTCAAGACTTATAAATACAACTCGA	572
Qy	606	ACGTCAATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTCAACGCGCACCGC	664
Db	573	GAACAAACCGGAGCTCCGGTTATGCTTAACGCTTTCGCGATTCAACCGGTAAACGGC	632
Qy	665	--GTATTGCGCGACTCATGCTATTACGCGCAGAGCTGCAGCGGTATAGTCACCGCGC	722
Db	633	GTACGTAGCTGTGGAGCTTTACGGCAGCTAGGTTTAAACGCGACAGTATATCGGT	692
Qy	723	ACACGAAACCGGTGTAGCTAAACCGCGTTTGGTCTATTTCATTTGGCGGTTATTGGCTAT	782
Db	693	TACCGCATTTGGTTAGCTAATAACAAGTCTAGCTTTCTTCAGTAAACCGTCTCTTTGCTTT	752
Qy	783	GTCCGAGATGATTTTACCTTACCAAGTTCAAGTACATCCCAATGGAGATTAAACACCGT	842
Db	753	AGGTGAATCTGATTTACCTTACCGCGTCCGATTAAACCGAATCAGGAGATTATGAAACGAT	812
Qy	843	TGGTCGGTTTCGATTTTGTATGACAAATAGAAATCCCAATGATTGCCACCGGAAAGTCGA	902
Db	813	CGGACGGTACGATTTTCGCGGAAATTTAGCGATGAGTATGACAGCTCATCTAAACCGGA	872
Qy	903	CCCGAATCCGGTGAATCTTCGCTTTAAGTACGAGCTGTTTCAAGCTTACCTTAAAC	962
Db	873	TCCAATAACCGGAGAAATTTTCGCTTTTCGCGTACCGTCCCGTTTCCA---CCGTTTTAAC	929
Qy	963	ATATCTCCGATTCACCGGACGAACTAAATACCGGAGCTTCAGAG---TTCAGCTTGA	1019
Db	930	ATATTTCCGTTTGTATTTCCGCGGAAACAAAGAGACGTTCCGATATTCTCGATGAC	989
Qy	1020	TCAGCCAAACGATGATGACGATTTTCGCGATTACAGAGAACTTTCGTCGTACCTGACCA	1079
Db	990	GTCTCCGTCGTTTCTCCATGATCTTCGCGATCACGAAACGTCACCGGATTTTCGAGAGAT	1049
Qy	1080	GCAAGTCG-----TTTTCAAGCTCCCGAGATGATCCGCGTGGGTCTCCGGTGGT	1130
Db	1050	TCAGCTTGGATGAGGATGAACATGTTGGATTGTTCTCGAAGGTGGTCTCCGGTTGG	1109

QY 1131 TTACGACAAAGCAAGTGGCGAATTCGGGATTTTAGACAAATAGCGGAAGATTCATC 1190
 Db 1110 TACTGATACCGGAAACTCCAGGCTTGGAGTATTCCTAAGTAGCGCGAGATCAGTC 1169
 QY 1191 GAAATTAAGTGGATGATGCTCAGATTCGATTCCTGCTCCATCTCTGGAACTCTGGGA 1250
 Db 1170 GGAGATGAATGGTTCGAAGTCTCTGGATTCATATCAATTCACGCTATTAATGCTTGGGA 1229
 QY 1251 AGAGCCAGAAACAGATGAAGTCGTCGTGATAGGTCCTGTATGACTCCACGAGCTCAAT 1310
 Db 1230 TGAAGATGATGAAACAGCGTCGTTTGAATGCACCGAATATTATGTCGATTGAACATAC 1289
 QY 1311 TTTCAAGAGTCTCAGAGATCTCAAGAGTGTCTGTCTGAATCGGCTGATCTCAA 1370
 Db 1290 TTTAGAGAGTGTATCTGGTTC---ATGCTTTGGTGGAGAGGTGAAGATCATCTGT 1346
 QY 1371 AACCGGTGAATCAACTCGCTCGATCATCTCCACAGAGATCAACAAGTCAACCTCGA 1430
 Db 1347 CACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391
 QY 1431 AGCAGGATGCTCAACAGAAACATGCTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTT 1490
 Db 1392 TTTGCTGCTGATTAATCCGGGTTTCTCGGAGATGTAGCAGTACGTTTACCGCGCAT 1451
 QY 1491 AGCCGAGCGGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAGT 1550
 Db 1452 TGGAGATCCGATGCCGAGATCTCCGGTGTGGTGAAGCTTGATGTCTAAAGGAGATCG 1511
 QY 1551 TAAGAAATC-----TTACGCGGATAACCGTACGAGGAGGCTCT 1595
 Db 1512 GGATGATGTAGGTGGCCCTAGATGTACGGTTCAGGTTGTTACGGCGGAGAACCGTT 1571
 QY 1596 GTTCTCCCGGAGAGAGGA-----GAGGAAGACGAAAGGATACATCCTCTG 1643
 Db 1572 TTTGCTAGTAGGATCTCGTAATCCGAGGCGGAGGAGATGATGTTATGTGTGAC 1631
 QY 1644 TTTGTTCCAGCAGAGAGATCGAATCGGATTCAGATAGTTAAGCCGTTAGCTT 1703
 Db 1632 GTATGTTCCAGATGAAGTGAATCGAGATCGAAGTTTCTGGTATGAGCGCTAAATCGCG 1691
 QY 1704 AGAGTTGAA-----GCAACGGTTAAACTTCGCTCAAGGTTCCGTAGCGATTTACCGG 1757
 Db 1692 GGAGCTTGAATCGTCGCGCGCTGAGGTTGCCGCGAAGGTTCCGTACGAGATTCATCG 1751
 QY 1758 TACATTCATCGGACCGGATGATTTGGCGAAGC 1789
 Db 1752 GTATTGTTCAAGGAAGTAGCTTAATAAGC 1783

RESULT 15

AAF77206
 ID AAF77206 standard; cDNA, 1950 BP.

AC AAF77206;

XX

XX 16-MAY-2001 (first entry)

XX cDNA encoding sunflower neoxanthin cleavage enzyme (NCE).

XX Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;

XX NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;

XX pathogen resistance; abscisic acid metabolism; ss.

XX Helianthus annuus.

XX Key

XX Location/Qualifiers

XX 1..1632

XX /tag= a

XX /product= "NCE"

XX /note= "Neoxanthin cleavage enzyme"

XX /partial

WO200112801-A2.

XX 22-FEB-2001.
 PD 17-AUG-2000; 2000WO-US22961.
 XX 18-AUG-1999; 99US-0149656.
 PF 23-MAY-2000; 2000US-0206405.
 PR (PION-) PIONEER HI-BRED INT INC.
 XX (CURA-) CURAGEN CORP.
 PA Bidney DL, Crasta OR, Hu X, Lu G;
 XX WPI; 2001-211215/21.
 PI P-PSDB; AAB72303.

Novel isolated defence-related signalling gene isolated from sunflower encoding neoxanthin cleavage enzyme, amino acid permease or glutamic acid-rich protein useful for increasing resistance of plant to a pathogen

Claim 1; Page 94-97; 135pp; English.

This invention relates to defence-related signalling genes isolated from the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich protein (GRP). The signalling gene is useful for increasing the resistance of a plant to a pathogen such as fungus, virus, bacterium, nematode or insect (e.g. European corn bore), preferably incorporating a *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a construct containing the gene into the genome of the plant. The gene is useful for regulating gene expression in a plant, in response to a stimulus such as infection with a pathogen, damage from a pathogen, hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid, oxalic acid or expression of a gene encoding oxalic acid oxidase. The genes are also useful for stem-preferred regulation of gene expression in a plant. The genes are useful in agriculture, particularly in the breeding of crop plants with improved agronomic traits, for modifying abscisic acid (ABA) metabolism and for modifying amino acid transport and content in plants. The present sequence represents cDNA encoding the sunflower neoxanthin cleavage enzyme (NCE).

Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Query Match 9.4%; Score 169.2; DB 22; Length 1950;

Best Local Similarity 51.0%; Pred. No. 18-43;

Matches 480; Conservative 0; Mismatches 453; Indels 9; Gaps 3;

QY 347 CGGCTGATCCTAGTGTTCAGATCGCGGAAATTTTGTCTCCGTTGAATGAACAGCCGTC 406

Db 308 CAGTTGATCCAAACACAGTTTGTCTGATACTTTTACCCGGTGGACGAATCCCTCCGA 367

QY 407 GGCGTAATCTTCGGTGTTCGGAATTCCTCCGATTCATCAAGAGTGTATGCGCA 466

Db 368 CTGACTGTGAAGTCATCGAGGCGCACTGCCAAGTTGCCCTTACTTCCGTA 427

QY 467 ACGGAGCTAACCCACTTCACGAGCCGGTGACAGGTACCACATTTCTTCGACGAGACGTA 526

Db 428 ATGTCGGAACCCGCAATTCCTTCGCGAGGACCCCTACACCTCTTCGATGCGATGCA 487

QY 527 TGGTTCACGCGCAATTCGHAACGGTTCAGCTAGCTACGCTTCCGCTTTACTCAGA 586

Db 488 TGCTCCATGCTATTCGTAICTCAATGGAAGTTCGTTATGTAGCCGATACATCAAA 547

QY 587 CTAAACCGGTTTGTTCAGGAACGTCATTCGGTTCGACCGGTTTCCCAAGCATCGGTG 646

Db 548 CATACAAATATTCATAGAGAAAGAGCGGATTCCTCCATTTATTCAAACGTTTTCAG 607

QY 647 AGCTTCACGCGCACACGGTATTGCCCGCTCATGCTATTC---TACGCCAGAGTGGAG 703

Db 608 GGTTAATGCTGACTGCTCTGAGCTCGCATGGCAGTCACTGCCGCGCATTTTGG 667

QY 704 CCGGTATAGTCGACCCCGCACACGGAAACCGGTGTAGCTAAACCGCGGTTTGTGCTATTCA 763

```
Db 668 CTGGACAATTGAGCCCAAAAGGATATGGCTCTAGCCAAATACCAGTCTGECCTTTTGG 727
Qy 764 ATGCGCGGTTATGGCTATGTCGGAGGATGATTTACCTTACCAAGTTACAGATCACTCCCA 823
Db 728 GCAACAGACTTTTGGCTCTGGAGAGTGGGATCTCCCATATGCGTCAAACTAGCGCCG 787
Qy 824 ATGGAGATTTAAACCGTTGGTGGTTCGATTTTGTATGGACAATTAGATCCACAATGA 883
Db 788 ACGGTGACATATGTCACCGTTCGGAGCTGAGGACTTCGACGGCAAACTATTATGAGCATGA 847
Qy 884 TTGCCCAACCCGAAAGTCCAGCCGGAATCCGGTGAACCTTCGCTTTAAGCTTACGACGTCG 943
Db 848 CCGCTCACCCCAAAATCGATCCAGTAGCAAGAGCTTTTGGCTTTTCGTTACGGTCCAG 907
Qy 944 TTTCAAAGCCTTACCTAAATACTTCCGATTTCTCAGCGAGCGAACTAAATCACCGGACG 1003
Db 908 TCCCC- --CCTTTCCTAACCTTTTTCGTTTCAACGAAACGGAGAAAAACAAGCCGATG 964
Qy 1004 TCGAGATTCAGC- --TTGATCAGCCACAGGATGATGCAGATTTCCGGAATTACAGAGAACT 1060
Db 965 TCCCGATCTTCTCAATGACAGCCGCTGTTTCTCCAGACTTCGCCATCACCAAAACT 1024
Qy 1061 TCGTCGTCGTACCTGACCAAGAGTCGTTTTCAGCTCCCGGAGATGATCCGCGTGGGT 1120
Db 1025 ACGCGATTTTCCCGAGATCCAAATCGGGATGAGCCCAATGGAGATGCTGGGTGGGGAT 1084
Qy 1121 CTCGGGTGTTTACGACAAGAACAGGTGCGAAGATTCGGGATTTTAGACAAATACGCCG 1180
Db 1085 CCCCAGTTAGCGCGAGCCTGGAAGGTGCTCGGCTCGGTTGATCCCTCGGTACGGGA 1144
Qy 1181 AAGATTCATCGAACHTTAAGTGGATTGATGCTCCAGATTGCTTCTGTTCCATCTCTGGA 1240
Db 1145 AAGACGAGTCCGAGATGAAGTGGTTGAGTTCCGGGTTTAAATGATGATACATTGCATCA 1204
Qy 1241 ACGTTGGAGAGCCAGAACAGATGAAGTCGTCGTGATAG 1282
Db 1205 ATGCATGGAGAGGATGCGGAGATACGGTGGTGTGTTGG 1246
```

Search completed: November 16, 2003, 15:31:21
Job time : 515 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2003, 20:16:12 ; Search time 3372 Seconds
(without alignments)
4317.432 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASTATAAASGRWLGCHT.....VPYGFHCTFIGADDLAKQVV 599

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cgn2_1/USPTO/spool/US09758269/runat 14112003 192309 25769/app query.fasta_1.775
-DE=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09758269@cgn 1 1 2810 @runat 14112003 192309 25769 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmul:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pg:*
27: em_gss_vri:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1256.5	39.9	837	28	BH662445	BH662445 BOHTU09TR
C 2	1108	35.2	805	12	BM408615	BM408615 EST582942
C 3	1073	34.1	766	29	BZ457573	BZ457573 BONGT80TF
C 4	1057	33.6	720	12	BM412731	BM412731 EST587047
C 5	1036	32.9	721	28	BH739063	BH739063 BOMHT72TF
C 6	1020	32.4	787	12	BM408565	BM408565 EST582892
C 7	1013	32.2	781	28	BH549344	BH549344 BONGRQ3TF
C 8	977	31.0	644	12	BM535408	BM535408 EST588430
C 9	956	30.3	643	10	AW933524	AW933524 EST59283
C 10	936.5	29.7	1884	11	AY106323	AY106323 Zea mays
C 11	909	28.9	637	13	BU550566	BU550566 GM880021A
C 12	888	28.2	667	29	AL950790	AL950790 Arabidops
C 13	876	27.8	592	10	BF113346	BF113346 EST440336
C 14	865	27.5	627	9	AV826228	AV826228 AV826228
C 15	854	27.1	668	29	CC157378	CC157378 ig19b12.b
C 16	847.5	26.9	594	28	BH458011	BH458011 BOHT02TF
C 17	846.5	26.9	696	28	BZ022957	BZ022957 oeh323c12
C 18	840	26.7	566	10	BZ461924	BZ461924 EST413439
C 19	826	26.2	564	12	BM085488	BM085488 saj37d09
C 20	825	26.1	553	12	BM085672	BM085672 saj38a02
C 21	821.5	26.1	711	29	BZ429431	BZ429431 BONDRL4TR
C 22	817	25.9	562	12	BM536135	BM536135 EST589157
C 23	815.5	25.9	549	12	BM891057	BM891057 sam22d05
C 24	815	25.9	617	10	BZ458861	BZ458861 EST414153
C 25	813	25.8	649	29	CC016051	CC016051 PUDGL12TD
C 26	804.5	25.5	547	12	B1974879	B1974879 sai74b11
C 27	802.5	25.5	602	13	BQ505126	BQ505126 EST612541
C 28	799.5	25.4	641	29	BZ525251	BZ525251 OGALK24TM
C 29	799	25.4	559	12	BM085005	BM085005 saj3la08
C 30	798	25.3	520	10	BZ432853	BZ432853 EST399478
C 31	798	25.3	547	10	BE437072	BE437072 EST408190
C 32	794	25.2	627	10	BF051297	BF051297 EST436472
C 33	791	25.1	657	9	AA556214	AA556214 69 Lob101
C 34	786.5	25.0	546	12	BM084948	BM084948 saj30b08
C 35	786	25.0	618	9	AW443298	AW443298 EST308228
C 36	769	24.4	578	29	BX004482	BX004482 Arabidops
C 37	765.5	24.3	612	29	BZ525247	BZ525247 OGALK24TC
C 38	761	24.2	532	12	BJ563195	BJ563195 BJ563195
C 39	759.5	24.1	567	13	BQ582721	BQ582721 E012280-0
C 40	749	23.8	648	29	BZ727312	BZ727312 OGPBF70TC
C 41	739	23.5	495	10	BE434930	BE434930 EST406008
C 42	734.5	23.3	509	10	BE451573	BE451573 EST402461
C 43	716.5	22.7	699	28	BH981407	BH981407 odel6e05
C 44	711	22.6	651	29	BZ462969	BZ462969 BONDND77TR
C 45	701.5	22.3	590	10	BG592987	BG592987 EST491665

ALIGNMENTS

RESULT 1
BH662445/c
LOCUS BOHTU09TR BO_2.3 KB Brassica oleracea genomic clone BOHTU09,
DEFINITION BOHTU09TR BO_2.3 KB Brassica oleracea genomic clone BOHTU09,
ACCESION BH662445
VERSION BH662445.1 GI:18721309
KEYWORDS genomic survey sequence.
SOURCE GSS.
ORGANISM Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.

```

REFERENCE 1 (bases 1 to 837)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
    Location/Qualifiers
        1..837
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHTU09"
            /clone_lib="BO 2.3 KB"
            /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"
BASE COUNT 171 a 238 c 220 g 208 t
ORIGIN
Alignment Scores:
Pred. No.: 3,78e-122 Length: 837
Score: 1256.50 Matches: 238
Percent Similarity: 93.19% Conservativeness: 22
Best Local Similarity: 85.30% Mismatches: 17
Query Match: 39.89% Indels: 2
DB: 28 Gaps: 1
US-09-758-269-6 (1-599) x BH662445 (1-837)
QY 270 ValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGln 289
DQ 836 GTCCAGATCACTCAAGGGAGACTGATAAAACCGTTGGCGGTACGATTTCACACACAGT 777
QY 290 LeuGluSerThrMetIleAlaHisProLysValAspProGluSer-GlyGluLeuPheAl 309
DQ 776 TTAGATCCACATGATCGCCACCCGAAAGTCGACCGGAATCCGGGGAGCTATTGCG 717
QY 309 aLeuSerTyAspValValSerLysProTyLeuLysTyPheArgPheSerProAspGln 329
DQ 716 TCTAAGCTACGACGTCTTCGAAGCCTTTTAAAGTACTTCAGACTCTACCGGACCG 657
QY 329 YThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAl 349
DQ 656 GATAAAATCAACCGGATGTCGAGATCAACCTCGATCAGCCGACGATGATGACGACTTCG 597
QY 349 alleThrGluAsnPheValValValProAspGlnValValPheLysLeuProGluMe 369
DQ 596 TATAACGGGAACCTCTGTGTATACCGGACCAAGTCTGTCTCAAGCTCCCGAGAT 537
QY 369 tIleArgGlyGlySerProValValTyAspLysAsnLysValAlaArgPheGlyIleLe 389
DQ 536 GATCCGCGGGCTCTCCGGTGTGTACGACAAAGGAGAAAGTTTCAAGATTCCGAGTTT 477
QY 389 uAspLysTyAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCy 409
DQ 476 AGACAAGTACCGCGGACACTCTGCGCCATCAGGTGATCGACCGCGGAGTGTCTTG 417
QY 409 sPheHisLeuTrpAsnAlaTrpGluLuproGluThrAspGluValValIleGlySe 429
DQ 416 CTTTCATCTCTGGAACCTTTGGGAGAGCCGGACACAGAGAGATCTGTGATCGGTC 357
QY 429 rCyMetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValle 449
DQ 356 GTGCATGACTCCCGCGGACTCAATTTCAACGAGCGCGAGAAATCTCGAGAGTCTCT 297
QY 449 userGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAs 469

```

```

Db 296 CTCCAAGATACGCTGAACCTCAGAACCGGAGAGTCCACTCGCGCTGATCATCTCGA 237
QY 469 nGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgly 489
DQ 236 CGGAGAGGAGCAAGTCAACCTCGAAGCAGGGATGTTTAAACCGAATCTCTCGGCGTAA 177
QY 489 sThrLysPheAlaTyLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLy 509
DQ 176 GACCAGTTTCGCTTACCTCGCTAGCGAGCGGTGGCCCTAAAGTCTCTGGCTTTGCGAA 117
QY 509 sValAspLeuThrThrGlyGluValLysLysHisLeuTyGlyAspAsnArgTyGlyCl 529
DQ 116 AGTTGATCTCTTTACCGGAGAGTCAAGAAAGTATCTGTACCGCGGTGACCGTTACGTGG 57
QY 529 YGluProLeuPheLeuProGly---GluGlyGlyGluGluAspGluGlyTyIle 546
DQ 56 AGAGCCTCTCTTTCTCCAGGAGACGAGAACAGAGCAGAGGACGACGGTTTACATC 2
RESULT 2
BM408615 805 bp mRNA linear EST 22-JAN-2002
LOCUS EST592942 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION cLEG45023 5' end, mRNA sequence.
ACCESSION BM408615
VERSION BM408615.1 GI:18260245
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 805)
AUTHORS Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Teal,
J., Bougri, O., Kirkness, E., Utterback, T., Van Aken, S., Renning,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES
    Location/Qualifiers
        1..805
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /cldes="cLEG45023"
            /tissue_type="Pericarp"
            /dev_stage="breaker"
            /lab_host="SOLR"
            /clone_lib="Tomato breaker fruit"
            /note="Vector: pBluescriptKCMUadapt; Site 1: EcoRI;
            Site 2: XhoI; supplier: Boyce Thompson Institute;
            Sequencing: The Institute for Genomic Research. Fruit
            were harvested at the breaker stage (first sign of
            lycopene accumulation on the blossom end of fruit). Fruit
            were cut in half and the seeds and locules were discarded
            prior to freezing the pericarp."
BASE COUNT 262 a 144 c 180 g 219 t
ORIGIN
Alignment Scores:
Pred. No.: 1.82e-106 Length: 805
Score: 1108.00 Matches: 212
Percent Similarity: 89.51% Conservativeness: 27
Best Local Similarity: 79.40% Mismatches: 26
Query Match: 35.17% Indels: 4

```

```

DB: 12 Gaps: 1
US-09-758-269-6 (1-599) x BM408615 (1-805)
QY 300 ValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyr 319
Db 5 CTCGACCCAGTTCGCGGTGAGCTATTGCTTAGCTACGATGCTGATTCAGAGCCATAC 64
QY 320 LeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluLeu 339
Db 65 CTCAGTACTTCAGATTTTCAAAAATGGGAAAATCAAAATGATGTTGAAAATCCAGTT 124
QY 340 AspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAsp 359
Db 125 GAAGACCCAAACATGATGATGATTCGCAATTACTGAGACTTCGTCGTCATTCCTGAT 184
QY 360 GlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAsp 379
Db 185 CAACAAGTCGTTTCAAGATGCTGAAATGATCCGTGGAGGTTCACCGGTGGTTACGAC 244
QY 380 LysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIle 399
Db 245 AAGAACAAAGTTTCCGATTTGGTATTCTGGTAAGTACGGGAAGATGGGTCTGATTTG 304
QY 400 LysTrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluPro 419
Db 305 AAATGGTGAAGTACCTGATTTCTGTTTCCACCTGCAATGCTTGGGAAGACA 364
QY 420 GluThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsn 439
Db 365 GAAACAGATGAATCGTTGTAATGGTTCATGATGACACCAACAGATTCATTTCAAT 424
QY 440 GluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGly 459
Db 425 GAATGTGATGAAGGCTAAAGAGTGTATTATCCGAATCCGCTCAATTTGAAACAGGG 484
QY 460 GluSerThrArgArgProIleLeuSerAsnGluAspGlnGlnValAsnLeuGluAlaGly 479
Db 485 AAATCAACAAAGAAATCCATATATCGAAACCCCGATGAACAGATGAAATTTAGAAGCTGA 544
QY 480 MetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGlu 499
Db 545 ATGGTGAACCGAAACAACTCGGAAGGAAACAGAGTATGCTTTATTTGGCTATCGCTGA 604
QY 500 ProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLys 519
Db 605 CCATGCCCAAAAGTTCTCGTTTTCGAAAGTAAACCTGTTCCACCGGTGAAGTTGAGAAA 664
QY 520 HisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlu 537
Db 665 TTCATTATGGTGACAAACATATGTTGGGGAACCTCTTTTTCACAAGA-GNACCCAAAC 723
QY 538 GlyGlyGluGluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys 557
Db 724 AGCCAGGAAGAAGACGATGTTATATTTAGCTTTTCGTTCCAGATGAGAAAGA-TGGGAA 782
QY 558 SerGluLeuGlnIleValAsn 564
Db 783 TCAAAACCTGCAATTTGGTAAC 803

RESULT 3
BZ457573/c
LOCUS BNGT80TF BO_1.6_2_KB tot Brassica oleracea genomic clone BONGT80,
DEFINITION genomic survey sequence.
ACCESSION BZ457573
VERSION BZ457573.1 GI:26737071
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.

```

REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 766)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of *Brassica oleracea*
 Unpublished
 Other GSSs: BONGT80TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES source

1..766
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone_lib="BO_1.6_2_KB tot"
 /clone="BONGT80"
 /note="Vector: pHO517 Site 1: BstXI; 1.6-2 kb sheared
 total DNA inserted into pHO51 using BstXI linkers"

BASE COUNT ORIGIN

177 a 203 c 203 g 183 t
 Alignment Scores:
 Pred. No.: 8 52e-103 Length: 766
 Score: 1073.00 Matches: 201
 Percent Similarity: 88.98% Conservative: 25
 Best Local Similarity: 79.13% Mismatches: 27
 Query Match: 34.06% Indels: 0
 DB: 29 Gaps: 0

US-09-758-269-6 (1-599) x BZ457573 (1-766)

QY 174 AspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPhe 193
 Db 759 GACGGTATGAT-CACCGGTTCGGTTTAAATACAGGTTCGGTTAGTTATTCGTCGCCGTTT 701
 QY 194 ThrGlnThrAsn-ArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAl 213
 Db 700 ACCGAACAACCCCGTTTGTTCACAGCGAGAGCTCGCGCGTCTGTTTCCCAAGC 641
 QY 213 alleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAl 233
 Db 640 GATAGGAGAGCTTCACGACACTTGGGGATCGCTAAGCTTATCTATTTAACCCACGTGG 581
 QY 233 aAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTy 253
 Db 580 GCTTTTCGGTTTACTGACCCGACCAAGGACTCGGTGTGGCTAATGCCGTTTAGCTTA 521
 QY 253 rPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleTh 273
 Db 520 TTTCAACGGTCATCTCTTAGCAATGTCGGAAGACGATTTACCGTACCGTAAAGTCAC 461
 QY 273 rProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerTh 293
 Db 460 CCAACCCGAGATTAGAACCTCGGCGCGGTACGATTTTCGAGGTTCAGTAAATTCGAC 401
 QY 293 rMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAs 313
 Db 400 AATGATCGCCCAACCCGAAACTCGACCCGGAACCCCGAGACTATTCGCTCTCAGTACGA 341
 QY 313 pValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerPr 333
 Db 340 CGTCGTTTCGAAGCTTACTTAAATACCTTCAGATTTCGATTCGAGGTTCGCAATCCCC 281
 QY 333 oAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGluAs 353
 Db 280 CGAGTCGAGATCCCGCTCGACCAACCGACGATGATCCAGACTTCGCGATCACCGAGAA 221

QY 353 nPheValValProAspGlnGlnValPheLysLeuProGluMetIleArgGlyG1 373
 Db 220 CTTTCGTCTGGTTCGGACCGACAGTGGTCTTTCAGGCTACCGAGATGATCCGCGCGG 161
 QY 373 ySerProValValTyAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyAl 393
 Db 160 CTCTCCGGTGGTTTACGACAAACAAAGATTCAGATTCGGTGTTCGAGCAATACGC 101
 QY 393 aGluAspSerSerAenIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTr 413
 Db 100 TGAAGACGCTTCGTGATCGCGGTGATCGAGTGAAGACTGCTTCGTTCCATCTCTG 41
 QY 413 pAenAlaTrpGluGluProGluThrAspGluValVal 426
 Db 40 GAACGCTTGGGAAGAGCGGACACAGACGAGGTCTGCTC 1

RESULT 4
 LOCUS BM412731 720 bp mRNA linear EST 22-JAN-2002
 DEFINITION EST587047 tomato breaker fruit Lycopersicon esculentum cDNA clone
 CLEGG60N24 5' end, mRNA sequence.
 ACCESSION BM412731
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Lycopersicon esculentum (tomato)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 720)
 REFERENCE
 AUTHORS Alcala, J., Vrbalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai,
 J.M., Fraser, C.M., Kirkness, E., Utterback, T., Van Aken, S., Ronning,
 C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from tomato fruit tissue, breaker stage (2002)
 Unpublished
 CONTACT: CUGI
 CLEMSON UNIVERSITY GENOMICS INSTITUTE
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 This clone is available through the Clemson University Genomics
 Institute
 Seq primer: T3.
 Location/Qualifiers
 1..720
 /organism="Lycopersicon esculentum"
 /mol_type="RNA"
 /cultivar="TA96"
 /db_xref="taxon:4081"
 /clone="CLEGG60N24"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /clone_lib="tomato breaker fruit"
 /note="vector: Bluescript-SKmuCudapt; Site 1: EcoR1;
 Site 2: XhoI; supplier: Boyce Thompson Institute. Fruit
 sequencing: The Institute for Genomic Research. Fruit
 were harvested at the breaker stage (first sign of
 lycopen accumulation on the blossom end of fruit). Fruit
 were cut in half and the seeds and locules were discarded
 prior to freezing the pericarp."

BASE COUNT 226 a 143 c 155 g 196 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,84e-101 Length: 720
 Score: 1057.00 Matches: 194
 Percent Similarity: 90.38% Conservative: 22
 Best Local Similarity: 81.17% Mismatches: 23
 Query Match: 33.56% Indels: 0
 DB: 12 Gaps: 0

US-09-758-269-6 (1-599) x BM412731 (1-720)

QY 248 AsnAlaGlyLeuValTyPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267
 Db 3 AACCGCGGTTTACTCTAATTAACCGAATTACTTGTCTGTAAGATGATTGGCT 62
 QY 268 TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp 287
 Db 63 TACCATTGAAGGTAAACCCACCGCGGATCTTAAACAGAGGGTCGATTCGATTTCCGAC 122
 QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
 Db 123 GGCCACGCTAAATCCACCATGATAGCTACCCAAAAGTCGACCCAGTTTCCGGTGAGCTA 182
 QY 308 PheAlaLeuSerTyAspValValSerLysProTyLeuLysTyPheArgPheSerPro 327
 Db 183 TTTGCTCTTAGCTACGATGTGATTCGAAGCCATACCTCAAGTACTTCAGATTTTCARAA 242
 QY 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAsp 347
 Db 243 AATGGGGAATAATCAANTGATGTGAAATTCAGTTGAAGACCCCAACATGATGATGAT 302
 QY 348 PheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuPro 367
 Db 303 TTCGCAATTACTGAGAACTTCGTCGTCATTCCTGATCAACAGTCGTTTTCAGATGCT 362
 QY 368 GluMetIleArgGlyGlySerProValValTyAspLysAsnLysValAlaArgPheGly 387
 Db 363 GAATGATCCGTGAGGTTTCACCGGTGGTTTACGACAGACAAAGTTTCCCGATTGGT 422
 QY 388 IleLeuAspLysTyAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCys 407
 Db 423 ATTCTGGATAAGTACCGAAGATGGGTCTGATTTGAAATGGGTGGAAGTACCTGATTGT 482
 QY 408 PheCysPheHisLeuTrpAenAlaTrpGluGluProGluThrAspGluValValIle 427
 Db 483 TTCTGTTTCCACTCTGGATGCTTGGGAAGACGAGAAACACATGAATCGTTGTAAT 542
 QY 428 GlySerCysMetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSer 447
 Db 543 GGTTTCATGTATGACACCCAGACTCCATTTCAATGAATGTGATGAAGGGCTAAAGAT 602
 QY 448 ValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIle 467
 Db 603 GTTTTATCCGAATCCGTCCTCAATTTGAAACAGGGAATTCACAGAAAATCCATAATC 662
 QY 468 SerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeu 486
 Db 663 GAAACCCGGATGAACCAAGTGGATTTAGAACTGGATGGTGAACCCGAAACAACTC 719

RESULT 5
 LOCUS BH739063 721 bp DNA linear GSS 20-FEB-2002
 DEFINITION BOMHT72TF BO_2_3 KB Brassica oleracea genomic clone BOMHT72,
 genomic survey sequence.
 ACCESSION BH739063
 VERSION BH739063.1 GI:18844458
 KEYWORDS
 SOURCE
 ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; euroids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 721)
 REFERENCE
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished
 COMMENT Other_GSSs: BOMHT72TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

FEATURES

Location/Qualifiers

1..721

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOMHT72"

/clone.lib="BO 2.3 KB"

/note="vector: pHO31, Site 1: BstXI; 2-3 kb sheared"

genomic DNA inserted into pHO31 using BstXI linkers"

188 a 178 c 198 g 157 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 6.4e-99 Length: 721

Score: 1036.00 Matches: 193

Percent Similarity: 91.56% Conservative: 24

Best Local Similarity: 81.43% Mismatches: 16

Query Match: 32.89% Indels: 4

DB: 28 Gaps: 2

US-09-758-269-6 (1-599) x BH739063 (1-721)

QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys 321

Db 4 CCGGAACCCCGAGCTATTGCCTCAGCTACGACGTCGTTTCGAAGCCTTACTTAAAA 63

QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluLeuLeuAspGln 341

Db 64 TACTTCAGATTCGCATCGACGCGTCCGAAATCCCGACGTCGAGATCCCGCTCGACAA 123

QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361

Db 124 CCGAGATGATCCACGATTCGCGATCCGACGAGAACCTTCGCGGTTCGCGACGACAA 183

QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLysAsn 381

Db 184 GTGGTGTTCAGGCTACCGAGATGATCGCGCGGCTCTCCGGTGGTTTACGACAAAC 243

QY 382 LysValAlaArgPheGlyLeuAspLysTyrAlaGluAspSerSerAsnIleLysTrp 401

Db 244 AAGAATCAAGATTCGGTGTGTTGAGCAATACGCTGAAGACGCTTCGTCATCGGTGG 303

QY 402 IleAspAlaProAspCysPheHisLeuTyrAsnAlaTyrGluLysProGluThr 421

Db 304 ATCGAGTGGAGATGCTTCTGTTTCCATCTCTGGAACGCTTGGGAAGGCGGAGACA 363

QY 422 AspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGluSer 441

Db 364 GACGAGGTGCTGCTATCGGTCATGATGACGCGCGGCTCGATATTCACGACAC 423

QY 442 AspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSer 461

Db 424 GACGAAACACTTCAGAGTGTGTTGCGAGATTAAGCTTAACCTGAAACAGGGAGTCC 483

QY 462 ThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481

Db 484 ACGCGTGGCGGTTATCTCC-----GAGAAATTAATCTCGAAGCGGTATGGTA 534

QY 482 AsnArgAsnMetLeuGlyArgGlyThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501

Db 535 ACCCGGAATCTTTTAGGTAGAAAAACGGGTTCGCTTACCTTGCCTTTAACGACCGGTGG 594

QY 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHisLeu 521

Db 595 CCTAAGTGTCCGGTTTCGGAAGTGGACTTATCCCGGAGAGAGTTTCAAAATATATA 654

QY 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGly 538

Db 655 TAGCGTGACGTAAGTCGGAGAGAGCTCTGGTTATGCCG---GAAGGG 702

RESULT 6

BM408565

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..787

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="CLEG45E23"

/tissue_type="Pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/note="Vector: pBluescriptSKmCvAdapt; Site 1: EcoRI;

Site 2: XhoI; supplier: Boyce Thompson Institute;

sequencing: The Institute for Genomic Research. Fruit

were harvested at the breaker stage (first sign of

lycopene accumulation on the blossom end of fruit). Fruit

were cut in half and the seeds and locules were discarded

prior to freezing the pericarp."

BASE COUNT 216 a 194 c 165 g 212 t

ORIGIN

Alignment Scores:

Pred. No.: 3.58e-97 Length: 787

Score: 1020.00 Matches: 193

Percent Similarity: 82.02% Conservative: 26

Best Local Similarity: 72.28% Mismatches: 40

Query Match: 32.38% Indels: 8

DB: 12 Gaps: 1

US-09-758-269-6 (1-599) x BM408565 (1-787)

QY 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGluSerAsn 82

Db 11 CCAAAACAGAAAAACAACAAC-----TCCTCTTCTTCTTCA 46

QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaAlaGluGly 102

Db 47 ACTTCCAAAGTGGAAATTTAGTGCAAGAACGAGCAATGCTTTAGTGTAGAAAGT 106

QY 103 PheLeuValSerHisGlyLysLeuHisProLeuProLysThrAlaAspProSerValGln 122

Db 107 GCTTTAACTAAACATCACTTGAACACCTTTGCCGAAAAACACGCCACCGAGTCCAG 166

QY 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142

Db 123 ILAALGCLYASNpheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142

```

Db      167 ATTTCTGGGAATTTTGTCCGGTACCGGAAATCCAGTCTGTCAATCTCTTCGGGTACC 226
Qy      143 GlyLeuProAspSerLeuValThrValArgAsnGlyValAsnProLeuHis 162
Db      227 GGAAATAATCCCAATGTCTTCAAGCGGTTTACGTTCGAACCGGAGCTTAACCCCTTTT 286
Qy      163 GluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValysPhe 182
Db      287 GMAACCAACCGCGGACACCATTTCTCGACGCGGAGCGGTATGTTTCACGCGGTCAATTC 346
Qy      183 GluHisGlySerAlaSerTyraCysArgPheThrGlnThrAsnArgPheValGlnGlu 202
Db      347 AAAATGGGTGGCTAGTTACCGCTTTTACCTGAAACAGAGAGGCTTGTTCAGAA 406
Qy      203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGly 222
Db      407 AAAGCTTTGGTTCGCTCTTTCCCTAAAGCCATTTGTTGAATTTACATGTTCACTCTGGA 466
Qy      223 IleAlaArgLeuMetLeuPheTyraAlaArgAlaAlaGlyIleValAspProAlaHis 242
Db      467 ATTGCAAGGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGACTTGTGATCACAGTAAA 526
Qy      243 GlyThrGlyValAlaAsnAlaGlyLeuValTyraPheAsnGlyArgLeuLeuAlaMetSer 262
Db      527 GGAAGCTGTTGCTGCAACCGCGGTTTAGTCTATTTCATTAACCGATTACTTGTATGCT 586
Qy      263 GluAspAspLeuProTyraGlnValGlnIleThrProAsnGlyAspLeuValThrValGly 282
Db      587 GAAGATGATTGCTTACCATGTAAAGTAAACACCCCGCGGCTCTTAAACACAGGGGT 646
Qy      283 ArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspPro 302
Db      647 CGATTGATTTCCAGCGGCGAGCTAAATTCACCATGATAGTCAACCAAGCTCGACCA 706
Qy      303 GluSerGlyLeuLeuPheAlaLeuSerTyraValValSerTyraValValSerTyraVal 322
Db      707 GTTTCGGTGAGCTATTGCTTAGCTAGTACGATGATGATTGATGATGATGATGATGATG 766
Qy      323 PheArgPheSerProAspGly 329
Db      767 TTCAGATTTTCAAAAATGGGG 787

RESULT 7
BH549344      781 bp DNA linear GSS 14-DEC-2001
LOCUS        BOGRQ53TF BOGR Brassica oleracea genomic clone BOGRQ53, genomic
DEFINITION   survey sequence.
ACCESSION    BH549344
VERSION      BH549344.1 GI:17801124
KEYWORDS     GSS.
SOURCE       Brassica oleracea
ORGANISM     Brassica oleracea
REFERENCE    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
              ; eurosid II; Brassicales; Brassicaceae; Brassica.
              1 (bases 1 to 781)
              Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              Unpublished
              Other GSSs: BOGRQ53TR
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TF
              Class: sheared ends.
              Location/Qualifiers
                1..781
                  /organism="Brassica oleracea"
                  /mol_type="genomic DNA"

```

```

/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGRQ53"
/clone_lib="BOGR"
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
       genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT 204 a 203 c 192 g 182 t
ORIGIN
Alignment Scores:
Pred. No.: 1.95e-96 Length: 781
Score: 1013.00 Matches: 190
Percent Similarity: 84.23% Conservative: 29
Best Local Similarity: 73.08% Mismatches: 41
Query Match: 32.16% Indels: 0
DB: 28 Gaps: 0
US-09-758-269-6 (1-599) x BH549344 (1-781)
Qy      92 AlaAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHis 111
Db      1 GCAGCGCGGATTCGGATTCGATGTCGCTGAGCGCGTTGATCTCAGCTGAGCAAGATACT 60
Qy      112 ProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsn 131
Db      61 CCTCTTCCCAAAACCGCTGATCCAGTATTCAAATCCGCGGAATATTATTCGCGGTGCGG 120
Qy      132 GluGlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGly 151
Db      121 GAATTTCCGCTCGAGAGTCTTCGACGTTGAAGGAACAAATCCCTGACTGCATTACCGGA 180
Qy      152 ValTyraValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePhe 171
Db      181 GTTATGTCCTAACGCTGCAATCCGATGTTCCAGCAATCTCGTCGATCATCTTTGTTTC 240
Qy      172 AspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyraAlaCys 191
Db      241 GACGAGACCGGAATGGTTCACGAGTTAAATACCAACCGGTTCCAGCAACGCTACGATGTC 300
Qy      192 ArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhePro 211
Db      301 CGGTTTACTAAACCGAGAGATTAATTCAGAAAAACGATTGGTTCGACCGGTTTTCCG 360
Qy      212 LysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyraAla 231
Db      361 AAAGCAATTCGCGAGCTTCACGCCACTCGGGAATCGACGTCGTCTGATGCTTTTTTACGCA 420
Qy      232 ArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeu 251
Db      421 CGTGGGCTTTCCGCGCTCGTCAACATCAAAACGCGCTCGGGTAGCGAACGCGGTTTG 480
Qy      252 ValTyraPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyraGlnValGln 271
Db      481 GTTTACTTCAATAACCGGCTTTTAGCAATGTTCAGAACGATTTACCGTACCAATTAATAA 540
Qy      272 IleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu 291
Db      541 ATAATCAACCGGCGACCTCCAAACGTTGGGCTTACGATTTCCGCGGTCACTATAAA 600
Qy      292 SerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSer 311
Db      601 TCATCGATGATCGCCACCCGAAACTCGACCCGCTTACAAAGAGAGCTCCACGCGCTAAGC 660
Qy      312 TyrAspValValSerLysProTyraLysTyraPheArgPheSerProAspGlyThrLys 331
Db      661 TACGACGCTGTTAAAGACCTTACCTGAAATATTTCAGATTTCTCGCGGAGCGGTTAAA 720
Qy      332 SerProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThr 351
Db      721 TCACCGGAGCTGAGATCCCGCTCGAGACTCCGACGATGGTTCAGATTTTCGCTATAACG 780

```

RESULT 8
BMS35408

```

LOCUS      BM535408                644 bp    mRNA    linear    EST 20-FEB-2002
DEFINITION EST588430 tomato breaker fruit Lycopersicon esculentum cDNA clone
cLEGE5N24 5' end, mRNA sequence.
ACCESSION  BM535408
VERSION     BM535408.1   GI:18814035
KEYWORDS   EST.
SOURCE      Lycopersicon esculentum (tomato)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1 (bases 1 to 644)
AUTHORS     Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning
C.M., Frazer,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished
COMMENT      Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES             Location/Qualifiers
     1..644
     /organism="Lycopersicon esculentum"
     /mol_type="mRNA"
     /cultivar="TA496"
     /db_xref="taxon:4081"
     /clone="cLEGE5N24"
     /tissue_type="Pericarp"
     /dev_stage="breaker"
     /lab_host="SOLR"
     /clone_lib="tomato breaker fruit"
     /note="Vector: pBluescriptSKmCtadapt; Site 1: EcoRI;
Site 2: XhoI; Supplier: Boyce Thompson Institute; Fruit
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT  195 a 129 c 138 g 182 t
ORIGIN
Alignment Scores:
Pred. No.:      9,44e-93      Length:      644
Score:          977.00      Matches:     178
Percent Similarity: 91.59%      Conservative: 18
Best Local Similarity: 83.18%      Mismatches:  18
Query Match:    31.02%      Indels:      0
DB:             12      Gaps:      0

US-09-758-269-6 (1-599) x BM535408 (1-644)

Qy      248  AsnAlaGlyLeuValTyrPheAnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267
Db      3   AACGCCGGTTAGCTATTTCATAAACCGATTACTTGTCTATGCTGAGATGATTGGCT 62
Qy      268  TyrGlnValGlnIleThrProAnGlyAspLeuLeuThrValGlyArgPheAsp 287
Db      63  TACCATGTAAGGTAACACCCACCGCGATCTTAAACACAGAGGGTTCGATTTCGAC 122
Qy      288  GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
Db      123  GGCACGTAATATCCACCATGATAGCTACCCAAAGCTCGACCCAGTTCCGGTGAGCTA 182
Qy      308  PheAlaLeuSerTyrAspValValserLysProTyrLeuLysTyrPheArgPheSerPro 327
Db      183  TTTCCTTAGCTACGATGTGATTCAGAGCCATACCTCAAGTACTTCAGATTTCAAA 242
Qy      328  AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAsp 347

```

```

Db      243  ATGGGGAATAATCAATGATGTTGNAATTCAGTTGAAGACCCACATGATGATGAT 302
Qy      348  PheAlaIleThrGluAsnPheValValValProAspGlnGlnValValPheLysLeuPro 367
Db      303  TTCCCAATTACTGAGAACTTCGTCGTCATTCCTGATCAACAAGTCGTTTTCAGATGCT 362
Qy      368  GluMetIleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGly 387
Db      363  GAAATGATCGTGGAGTTCCCGGTGGTTTACGACAGAACAAAGATTTCGCCGATTGGT 422
Qy      388  IleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTyrIleAspAlaProAspCys 407
Db      423  ATTCTCGATAAGTACGCGAAAGATGGTCTGATTGAAATGGGTTCAGATCACTGATTGT 482
Qy      408  PheCysPheHisLeuThrAsnAlaTyrGluGluProGluThrAspGluValValIle 427
Db      483  TTCTGTTTCCACCTCTGGAAATGCTTGGGAAGACGAGAACAGATGAATCTGTTGTAAT 542
Qy      428  GlySerCysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLysSer 447
Db      543  GGTTCATGATGACACCCAGACTCCATTTCAATGAATGTGATGAGAGGCTAAAGAGT 602
Qy      448  ValLeuSerGluIleArgLeuAsnLysThrGlyGluSer 461
Db      603  GTTTTATCGAAATCCGCTCTCAATTGAAACACGGGAATAAC 644

RESULT 9
AW933524                643 bp    mRNA    linear    EST 18-MAY-2001
LOCUS      EST39283 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF54E2 5', mRNA sequence.
ACCESSION  AW933524
VERSION     AW933524.2   GI:11388127
KEYWORDS   EST.
SOURCE      Lycopersicon esculentum (tomato)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.B., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Frazer,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished
COMMENT      On May 30, 2000 this sequence version replaced gi:8108841.
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES             Location/Qualifiers
     1..643
     /organism="Lycopersicon esculentum"
     /mol_type="mRNA"
     /cultivar="TA496"
     /db_xref="taxon:4081"
     /clone="cLEF54E2"
     /tissue_type="fruit pericarp"
     /dev_stage="mature green (3-5 days pre-ripening)"
     /lab_host="SOLR"
     /clone_lib="tomato fruit mature green, TAMU"
     /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed immature, and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT  184 a 130 c 142 g 186 t 1 others
ORIGIN

```

Alignment Scores: 1.57e-90 Length: 643
 Pred. No.: 956.00 Matches: 173
 Score: 956.00 Conservative: 19
 Percent Similarity: 90.14% Mismatches: 21
 Best Local Similarity: 81.22% Indels: 0
 Query Match: 30.35% Gaps: 0
 DB: 10

US-09-758-269-6 (1-599) x AW933524 (1-643)

QY 230 TyrAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAla 249
 Db 4 TAGCTCTGGGGCTCTTCGGACTTGTGTATCAGTAAGAGAACTGGTGTTCACACGCC 63

QY 250 GlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGln 269
 Db 64 GGTTAGTCTATTTCATAACCCGATTACTTGTCTGATGCTGAAGATGATTGCTTACCAT 123

QY 270 ValGlnIleThrProAsnGlyAspLeuValGlyArgPheAspPheAspGlyGln 289
 Db 124 GTAAGGTAACACCCACCGCGATCTTAAACAGAGGGTGTGATTCGATTCGACGCCAG 183

QY 290 LeuGluSerThrMetIleAlaHisProLysValAspProLysSerGlyGluLeuPheAla 309
 Db 184 CTAATATCCACCATGATGATGATCACCACCAAGCTCGACCCAGTTTCGGTGAGCTATTGCT 243

QY 310 LeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGly 329
 Db 244 CTAGCTACCATGATGATTCAGAGCCCATCTCAAGTACTTCAGATTTTCACAAAATGGG 303

QY 330 ThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAla 349
 Db 304 GAAAAATCAATGATGTTGAAATCCAGTTGAAGACCCCAACATGATGATGATTCGCA 363

QY 350 IleThrGluAsnPheValValProAspGlnValValPheLysLeuProGluMet 369
 Db 364 ATTAAGTACAGAACTTCGTCGTCATTCGATCAACAGTCGTTTCAAGATGCTGAAATG 423

QY 370 IleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeu 389
 Db 424 ATCCGTGGAGTTTACCGGTGGTTTACGACAGAACAAAGTTTCCCGATTTGATTCG 483

QY 390 AspLysTyrAlaGluAspSerSerAsnIleLysTrpLeuAspLapProAspCysPheCys 409
 Db 484 GATAGTACGCGAAGATGGGTCTGATTTGAATGGGTGATGATGATGATGATGATGATG 543

QY 410 PheHisLeuTrpAsnAlaTrpGluProGluThrAspGluValValIleGlySer 429
 Db 544 TTCACCTCTGGAATGCTTGGAGAGACGACAGACAGATGAAATCGTTGTAATGGTTCA 603

QY 430 CysMetThrProProAspSerIlePheAsnGluSerAsp 442
 Db 604 TGATGACACCAACCACTCCATTTTCAATGAATGTGAT 642

RESULT 10
 AY106323 1884 bp mRNA linear HTC 16-OCT-2002
 Locus Zea mays PC0084517 mRNA sequence.
 DEFINITION Zea mays PC0084517 mRNA sequence.
 ACCESSION AY106323
 VERSION AY106323.1 GI:21209401
 KEYWORDS HTC.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 1884)
 Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 Unpublished (2002)
 JOURNAL
 REFERENCE 2 (bases 1 to 1884)

Coe, E.H.
 Direct Submission
 Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
 www.zmdb.iastate.edu.
 Location/Qualifiers
 1. 1884
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeDB:634734"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed Dupont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"
 BASE COUNT 505 a 402 c 477 g 500 t
 ORIGIN

Alignment Scores: 8.62e-98 Length: 1884
 Pred. No.: 936.50 Matches: 211
 Score: 936.50 Conservative: 99
 Percent Similarity: 55.16% Mismatches: 195
 Best Local Similarity: 37.54% Indels: 57
 Query Match: 29.73% Gaps: 15
 DB: 11

US-09-758-269-6 (1-599) x AY106323 (1-1884)

QY 66 SerSerAsnSerProAlaIleValValProLysAlaLysGluSerAsnThrLysGln 85
 Db 94 AGCCACCAAGACGCGCGTGTGTGCGCGCGCGCGCGCGT----- 141

QY 86 MetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuVal 105
 Db 142 -----AAGGGCTGCTTCTCTGGCGCTCGACCTGCTTGTAGTCCCTCGCGGTG 189

QY 106 -----SerHisGluLysLeuHisProLysThrAlaAspProSerValGlnIle 123
 Db 190 CGCTCGCCACGACGACCAAGCGCTCCAC-----TGGCTC 228

QY 124 AlaGlyAsnPheAlaPro---ValAsnGluGlnProValArgArgAsnLeuProValVal 142
 Db 229 TCGGCACTTCGCGCGCGCTGCTCGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 288

QY 143 GlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHis 162
 Db 289 GCACACCTCCCGAGTGTGTAATGAGAGTGTGTCAGGGTTGGCGCTTAATCCGAAGTTT 348

QY 163 GluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLysPhe 182
 Db 349 GCTCTGTTGCGGGTATCATTGTTTGTATGAGAGCGGATGATTCATGCGCATGCGTATT 408

QY 183 GluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGlu 202
 Db 409 AAGGATGGAAGAACTACCTATGATCAAGATATGTGAAGACTGCCCGCTCAACCAAGAG 468

QY 203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGly 222
 Db 469 GAGTATTTGGTGGAGCAAGTTTATGAAG---ATTGGAGACCTTAAGGATTTTGGGA 525

QY 223 IleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAlaHis 242
 Db 526 TTGTTTATGTCACCAATGACCAACTTCGGAAGAAATTCAAAGTCTTGTGATTTTACCTAT 585

243 GlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSer 262
 |||||
 586 GGATTGGGACAGCTAATCTGCTTATATATCATCATGTTAACTCATGGCTGTGCA 645
 |||||
 263 GluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGly 282
 |||||
 646 GAAGCAGATAGCCATATGTTTGAAGTCTTGAAGATGAGACTTGCAGACTCTTTGGC 705
 |||||
 283 ArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspPro 302
 |||||
 706 TTGTTGATTAAGAAAGAGTGAACATTTCTTTTACTGCCATCCAAAGGTTGACCT 765
 |||||
 303 GluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerTyrProTyrLeuLysTyr 322
 |||||
 766 TTTACAGATGAATGTTTCACATCCGATATTC---CATGAACCTCCATCTGTACATAC 822
 |||||
 323 PheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnPro 342
 |||||
 823 CGTGTGATTAAAGAGGAGCTATGCTTGCATCTGTGCCAATAACAATACCGGAATCT 882
 |||||
 343 ThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGlnVal 362
 |||||
 883 GTAATGATGATGATTTTGCATCACAGAAATCTACTATTTTATGGACCTCCCTTTA 942
 |||||
 363 ValPheLysLeuProGluMetIleArgGlySerProValVal---TyrAspLysAsn 381
 |||||
 943 TTGTTCCACCAAGGAATGTTGAAGAACCGTGAGTTTATCTACAAGTTTGATCTCTACA 1002
 |||||
 382 LysValAlaAaGpPheGlyLeuAspLysTyrAlaGluAspSerAsnIleLysTyr 401
 |||||
 1003 AAGAAGCTGTTTGGTATTTCTCCCGCTATGCAAGGATGACAACTCATCAGATGG 1062
 |||||
 402 IleAspAlaProAspCysPheHisLeuThrPheAlaThrGluGluProGluThr 421
 |||||
 1063 TTTCAACTCCCTAATTGTTTCATATTCATATGCTATGCTTGGAA-----GAGGGT 1116
 |||||
 422 AspGluValValIleGlySerCysMetThrProProAsp-----Ser 436
 |||||
 1117 GATGAAGTTGTTCTCATTTACTCGCGCTTCCAGATCCAGATTTGGCAAGTGAATGGA 1176
 |||||
 437 IlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeu 456
 |||||
 1177 TATCAAGTGACACAGCTCGAAACTTCGGGAATGAGCTGTACGAGATGAGATTCACATG 1236
 |||||
 457 LysThrGlyGluSerThrArgArgProIleLeuSerAsnGluAspGlnGlnValAsnLeu 476
 |||||
 1237 AAAACGGTGTCTCTCAAAAG-----CAATTGCTGT 1272
 |||||
 477 GluAla-----GlyMetValAsnArgAsnMetLeuGlyAaGlyThrLysPheAla 493
 |||||
 1273 TCTGCTGTGGATTTTCTCGTGTAAATGAGAGCTATATCTGCGAGAAAGCAGCGGTATGTC 1332
 |||||
 494 TyrLeuAlaLeuAlaGluProThrProLysValSerGlyPheAlaLysValAsp----- 511
 |||||
 1333 TACTGCATATATCTTGACAGCATTCGGAAGGTGACTGTCATCAATAAGTTTGTGCTGCAT 1392
 |||||
 512 -----LeuThrThrGlyGluValLysLysHisLeuTyr 522
 |||||
 1393 GCTGAACCGGAAGTGGTGTGAAGAACTTGAAGTGGGAGGAATGTACAAAGCATATAT 1452
 |||||
 523 -----GlyAspAsnArgTyrGlyGlyGluProLeuPheLeu-----ProGlyGlu 537
 |||||
 1453 GACCTGGGACCTGGTAGATTTGGTTTCAGAGGCGATTTTGTTCACCAAGATCCAGGTGTG 1512
 |||||
 538 GlyGlyGluGluAspGlyTyrIleLeuCysPheValHisAspGluLysThrTriLys 557
 |||||
 1513 TCCGGAAGAAGATGACCGCTATTGATATCTTTTGTACAGCAGCAACACAGGGA 1572
 |||||
 558 SerGluLeuGlnIleValAsnAlaValSerLeuGluValGlu-----AlaThrValLys 575
 |||||
 1573 TCTGAAGTAATCTTATCATGATCAAGACATGCTGTGTCATCCAGTTGCGGTGTTGAG 1632

QY 576 LeuProSerArgValProTyrGlyPheHisGlyThrPheIleGlyAlaAspLeuAla 595
 |||||
 1633 CTTCTAATAGGTTCTTATGATTCATCCCTTTTGTAACTGAGCACTGGCT 1692
 |||||
 QY 596 LysGln 597
 |||||
 Db 1693 CGACAG 1698
 |||||
 RESULT 11
 LOCUS BU550566/c
 DEFINITION GM880021A20F06 Gm-r1088 Glycine max cDNA clone Gm-r1088-7932 3',
 mRNA sequence.
 ACCESSION BU550566
 VERSION BU550566.1 GI:22933427
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 637)
 REFERENCE Vodka, L., Shoemaker, R., Keim, P., Retzel, E., Khanna, A., Shealy, R.,
 Clough, S., Thibaud-Nissen, F., Corvelli, V., Erpelid, J., Raph, C.,
 Shoop, E., Stromvik, M., Schweitzer, P., Gong, G., and Liu, L.
 A Functional Genomics Program for Soybean (NSF 9872565) (2002)
 Unpublished
 OTHER ESTS: B7893117 corresponding to Gm-cl068-3283 (5')
 Contact: Vodka, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Tel: (217) 333-4582
 Fax: (217) 333-4582
 Email: l-vodka@uiuc.edu
 Insert Length: 637 Std Error: 0.00
 Plate: GM880021A20 row: F column: 06
 Seq primer: 5'-TTTCTTTTCTTTTCTTTTCTTTT(A/C/G)-3'
 High quality sequence stop: 637.
 Location/Qualifiers
 1..637
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone_lib="Gm-r1088-7932"
 /clone_lib="Gm-r1088"

FEATURES
source

/note="The library Gm-r1088 is a sequence-driven, reracked
 set of 9,216 low redundancy clones selected from cDNA
 libraries from various tissues and stages of development
 of soybean that consists of 2,706 cDNAs from germinating
 cotyledons (source library Gm-cl027); 1,355 cDNAs from
 immature seed coats (libraries Gm-cl019 and Gm-cl023); 917
 cDNAs from tissue culture derived somatic embryos (source
 libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from
 germinating seedlings, shoot tips, or leaves exposed to
 various stresses (source libraries Gm-cl064, Gm-cl065,
 Gm-cl066, and Gm-cl067; and Gm-cl068); and 963 cDNAs from
 young leaves exposed to bacterial and fungal pathogens
 (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The
 5' ESTs of the source clones from the different libraries
 was used to select singletons, or a representative of each
 contig, which were reracked to form library Gm-r1088 and
 the cDNA clones of the reracked Gm-r1088 library were then
 sequenced at the 3' end. The unigene selection and 3'
 sequencing was funded by NSF Plant Genome project #9872565
 (http://soybeanomics.cropsci.uiuc.edu/) as part of
 creation of a low redundancy soybean cDNA set. The source
 cDNA libraries were constructed by the laboratories of
 Lila Vodka, University of Illinois, Randy Shoemaker, Iowa
 State University, and Paul Keim, Northern Arizona

University as part of the Public EST project.
<http://129.186.26.94/soybeanest.html>. The contig analysis
 to select unique genes was performed by the laboratory of
 Ernest Retzel, Center for Computational Genomics and
 Bioinformatics, University of Minnesota,
<http://web.ahc.umn.edu/biodata/nfs/soy/>. Reracking and 3'
 sequencing were conducted by services of the University of
 Illinois Kerk Center for Comparative and Functional
 Genomics <http://www.biotech.uiuc.edu/keck.htm>. Note: The
 corresponding 5' EST from each clone in the Gm-r1088
 library is listed in the 'OTHER EST' field. The detailed
 information on the source library for each clone can also
 be obtained by referring to the Genome Systems clone ID of
 the original cDNA library that is also listed under 'OTHER
 EST'."

BASE COUNT 160 a 180 c 126 g 171 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,45e-85 Length: 637
 Score: 909.00 Matches: 168
 Percent Similarity: 92.38% Conservative: 26
 Best Local Similarity: 80.00% Mismatches: 14
 Query Match: 28.86% Indels: 2
 DB: 13 Gaps: 2

US-09-758-269-6 (1-599) x BUS50566 (1-637)

QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyAlaGluAspSerSerAsnIleLys 400
 DB 636 AACAGATTTCAGATTCGGATTCTCGAACAGATGCTAAAGATGCAATGATGAA 577
 QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu 420
 DB 576 TGGATCGATGCTCCCGAGTCTTCTGTTCCACTCTGGAACGCTGGAGAGCGGAA 517
 QY 421 ThrAspGluValValIleGlySerCysMetThrProAspSerIlePheAsnGlu 440
 DB 516 AATGATGAGATCTGTTGTCATCGTTCTGCATGACCCCGCGACTCCATTTTCAACGAA 457
 QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 DB 456 TGGCAAGAGAGTTTGAAGAGCATCTTCTCGAGATTAAGCTGAATTTGAAGCAGCGAAG 397
 QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
 DB 396 TCCACGAGAAACCCATTATCTCG---GAATCCGACACAGTCACTTGAACCGCGATG 340
 QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyLeuAlaLeuAlaGluPro 500
 DB 339 GTCAACAGAAACAAAGCTCGGAGAAACACGAAAGTTCGGTACTTAGCACTTCGGAGCG 280
 QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 DB 279 TGGCCTAAGGTTTCGGGTTTGCCAGGTTGATTTGTTTCACTGGGAGGTTAAGAGTAC 220
 QY 521 LeuTyArgLysAspAsnArgTyArgLysGlyGluProLeuPheLeuPro---GlyGluGly 539
 DB 219 ATGATATGCGAAGAGAGGTTTCGGAGGAGGAGGCTCTGTTTCTCCAAATGTTGTCAGCT 160
 QY 540 GluLysAspGluGlyTyrlleLeuLysPheValHisAspGluLysThrTrpLysSerGlu 559
 DB 159 GATGAAGACGATGCGCATATCTCGCATTCGTGATGATGAGAGGAGTGGAAATCGAG 100
 QY 560 LeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArg 579
 DB 99 CTGCAGATTGTCATGCGAAACACTTGAAGCTTCAGGCTTCAGTTAAGCTTCCTTCAGA 40
 QY 580 ValProTyArgLysPheHisGlyThrPheIle 589
 DB 39 GTTCTTATGGGTTTCACGGTACTTTTATT 10

RESULT 12

AL950790
 LOCUS
 DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-331D05-016046,
 genomic survey sequence.

ACCESSION AL950790.1 GI:24407412

VERSION GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
 and Weissshaar,B.
 A pipeline for automated high-throughput generation of ESTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines

TITLE

Unpublished

JOURNAL

REFERENCE

2 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics

TITLE

Unpublished

JOURNAL

REFERENCE

3 (bases 1 to 667)

TITLE

Unpublished

JOURNAL

REFERENCE

Direct Submission
 Rosso,M., Strizhov,N., Li,Y. and Weissshaar,B.
 Submitted (21-OCT-2002) Weissshaar B., Max-Planck-Institut fuer
 Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion close to or within gene Atg30100. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

1..667

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-331D05-016046"

/notes="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector pAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

BASE COUNT 176 a 145 c 161 g 184 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 2.57e-83 Length: 667
 Score: 888.00 Matches: 164
 Percent Similarity: 86.30% Conservative: 25
 Best Local Similarity: 74.89% Mismatches: 30
 Query Match: 28.19% Indels: 0
 DB: 29 Gaps: 0

US-09-758-269-6 (1-599) x AL950790 (1-667)

QY 171 PheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyAla 190

DB 7 TTCGACGGTACGGTATGTTTCACGCCCTTAAATCACTAACGGAGACGCGAGTACTCG 66

QY 191 CysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhe 210

DB 67 TCCCGGTTTACGGAACCGAGAGATGTTTCAAGAGAAACCACTCGTTCCTCCGATTTC 126

211 ProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyr 230
 Db CTAAGCTATAGGTAGCTACGTCTGCTGCAATCGACGATGATCTATTTC 186
 QY 231 AlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAspGly 250
 Db GCACGTGGTTTATTCGGTTTATTAATACAAAACGGAACGGAGTTGCTAAACGCCGT 246
 QY 251 LeuValTyrPheAsnGlyArgLeuAlaMetSerGluAspLeuProTyrGlnVal 270
 Db TTGGTTTACTTCACGACGGTTTATAGCTATGCTGAAGATGATCTACCTTACCAAGTT 306
 QY 271 GlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspGlyGlnLeu 290
 Db CGTGTCACCTGACAAATGGCGATTTAGAGACCATCGAAGATTTCGATTCGACGACACTA 366
 QY 291 GluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeu 310
 Db AGCTCCGCCATGATCGCTCACCGGAGATTGATTCGGTAACGAAGAGCTATTTCGGTTG 426
 QY 311 SerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThr 330
 Db AGCTACGACGCTGTTAAGAAACCGTATTGTAATACCTTTAAATTCCTCCGCGAAGGTGAG 486
 QY 331 LysSerProAspValGluLeuGlnLeuAspGlnProThrMetMetHisAspPheAlaIle 350
 Db AATACACCGGAGCTTGAGATTCTCTCGCCAGTCCGACGATGATCGAGATTTCGCGATC 546
 QY 351 ThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIle 370
 Db ACTGAGAATTCTGTTGATTCCGGATCAACAAGTTGTTTAACTCTCCGATATGTTT 606
 QY 371 ArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeu 389
 Db CTTGGGAATCTCCGGTTAAATACGACGAGAGAAATTTCCCGTTTGGAAATGTTG 663

RESULT 13
 BF113346
 LOCUS
 DEFINITION EST440936 tomato breaker fruit Lycopersicon esculentum cDNA clone
 CLEG44B9 5' sequence, mRNA sequence.
 ACCESSION BF113346
 VERSION BF113346.1 GI:10943036
 KEYWORDS EST.
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 Zuckeryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 592)
 AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Renning,C.M.,
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
 S.D.
 TITLE Generation of ESTs from tomato fruit tissue, breaker stage
 JOURNAL Unpublished
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1..592
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLEG44B9"
 /tissue_type="pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /clone_lib="tomato breaker fruit"

/notes="Vector: pBluescriptSKmCUadapt; Site 1: EcoRI;
 Site 2: XhoI; supplier: Boyce Thompson Institute;
 sequencing: The Institute for Genomic Research. Fruit
 were harvested at the breaker stage (first sign of
 lycopene accumulation on the blossom end of fruit). Fruit
 were cut in half and the seeds and locules were discarded
 prior to freezing the pericarp."

BASE COUNT 194 a 106 c 131 g 161 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4,02e-82 Length: 592
 Score: 876.00 Matches: 160
 Percent Similarity: 91.33% Conservatives: 19
 Best Local Similarity: 81.63% Mismatches: 17
 Query Match: 27.81% Indels: 0
 DB: 10 Gaps: 0

US-09-758-269-6 (1-599) x BF113346 (1-592)

QY 299 LysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysPro 318
 Db 3 AAGCTCGACCCAGCTTTCCGGTGAGCTATTGCTCTTAGCTACGATGATTCAGAACGCA 62
 QY 319 TyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluLeuGln 338
 Db 63 TACTCTAGTACTCTCAGATTTCAAAAATGGGAAAAATCAATGATGTTGAAATCCA 122
 QY 339 LeuAspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValPro 358
 Db 123 GTTGAAGACCCAAACAAATGATGATGATTTCCGCAATTAATCTAGCAATTCGTCGTCATCT 182
 QY 359 AspGlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyr 378
 Db 183 GATCAACAAGTCGTTTTCAGATGCTGAAATGATCCGTGGAGGTTCCCGTGGTTTAC 242
 QY 379 AspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsn 398
 Db 243 GACAAGAACAAAGATTCCCGATTGTTGTTAATGTTGATGATGACACCAACGACTCCATTTTC 302
 QY 399 IleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGlu 418
 Db 303 TTGAATGGGTTGAAGTACCTGATGTTTCTGTTTCCACTCTCGAATGCTTCGGAGAA 362
 QY 419 ProGluThrAspGluValValIleGlySerCysMetThrProAspSerIlePhe 438
 Db 363 GCAGAAACAGATGAATCGTTGTTAATGTTGTTTATGATGATGACACCAACGACTCCATTTTC 422
 QY 439 AsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThr 458
 Db 423 AATGAATGTCATGAAGCGCTAAAGAGTGTGTTTATCCGAAATCCGTCCTCAATTTGAAACA 482
 QY 459 GlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAla 478
 Db 483 GGGAAATCAACAGAAATCCATAATCGAAACCCGATGAACAAGTGAATTTAGAAGCT 542
 QY 479 GlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyr 494
 Db 543 GGAATGTTGAACCGAAACAAACTCGAAGGAAAAACAGATATGCTTAT 590

RESULT 14

AV826228

LOCUS

DEFINITION

AV826228 RAFL8 Arabidopsis thaliana cDNA clone

AV826228

ACCESSION

VERSION

AV826228.1

GI:19868288

KEYWORDS

EST.

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopses.

AV826228 627 bp mRNA linear EST 01-APR-2002
 LOCUS Arabidopsis thaliana cDNA clone RAFL08-11-H16 5',
 mRNA sequence.

AV826228

ACCESSION

VERSION

AV826228.1

GI:19868288

KEYWORDS

EST.

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopses.

```

REFERENCE
AUTHORS
1 (bases 1 to 627)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekic@cc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
TITLE
JOURNAL
COMMENT

FEATURES
source
1. .627
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF108-11-H16"
/dev_stage="rosette plants"
/lab_host="DH10B"
/clone_lib="RAF108"
/notes="Site 1: BamHI; Site 2: SalI; subjected to
dehydration-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT 162 a 179 c 128 g 158 t
ORIGIN

Alignment Scores:
Pred. No.: 6.37e-81 Length: 627
Score: 865.00 Matches: 168
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.46% Indels: 0
DB: 9 Gaps: 0

US-09-758-269-6 (1-599) x AV826228 (1-627)

Qy 1 MetAlaSerPheThrAlaThAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
Db 124 ATGGCTTCTTTCACGGCAACGGCTGCGGTTCTGGGAGATGGCTTGGTGGCAATCATACT 183
Qy 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 184 CAGCGCGCATATTCGTCTTCTCAAAGCTCCGACTTGAGTTATTGTAGTCTCTTACCTATG 243
Qy 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
Db 244 GCCAGTCGTGTACACGTAAGCTCAATGTTTCATCTCGGTTCCACATCTCCAGCTCTT 303
Qy 61 HisPheProLysGlnSerSerAsnSerProAlaLeuValLysProLysAlaLysGlu 80
Db 304 CATTTCCCTAAGCAATCATCAATCTCCCGCCATTGTTGTATAGCCCAAGCAAGAA 363
Qy 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
Db 364 TCCACACTAAACAGATGAATTTGTTCCAGAGAGCGGCGGCGAGCGTTGGACGGCG 423
Qy 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
Db 424 GAGGGTTTCTTGTTCAGCCACGAGAGCTACACCGCTTCCATAAACGGCTGATCTTAGT 483
Qy 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
Db 484 GTTCAGATCGCGGGAATTTTGTCTCCGGTGAATGAACAGCCGCTCCGGCGGTATCTTCG 543

Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaLeuPro 160
Db 544 GTGGTCGGAATAATCCCGATTCCATCAAGAGGTATGTCGCAACGAGCTACCCCA 603
Qy 161 LeuHisGluProValThrGlyHis 168
Db 604 CTTACAGAGCGGTGACAGGTAC 627

RESULT 15
CC157378 668 bp DNA linear GSS 29-APR-2003
LOCUS ig19b12.b1 WGS-Zmaysf (DH5a methyl filtered) Zea mays genomic clone
DEFINITION ig19b12, genomic survey sequence.
ACCESSION CC157378
VERSION CC157378.1 GI:30182158
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE
AUTHORS Rabinowicz, P., O'Shaughnessy, A.L., Ballija, V., Dedhia, N.,
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, I., McCombie, W.R., and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
1 (bases 1 to 668)
TITLE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
COMMENT clade; Panicoidae; Andropogoneae; Zea.
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: ig19 row: b column: 12
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 668.
Location/Qualifiers
1. .668
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ig19b12"
/lab_host="DH5a"
/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
/notes="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."
BASE COUNT 89 a 229 c 248 g 102 t
ORIGIN

Alignment Scores:
Pred. No.: 1.02e-79 Length: 668
Score: 854.00 Matches: 163
Percent Similarity: 83.33% Conservative: 22
Best Local Similarity: 73.42% Mismatches: 34
Query Match: 27.11% Indels: 3
DB: 29 Gaps: 1

US-09-758-269-6 (1-599) x CC157378 (1-668)

Qy 188 SerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArg 207
Db 3 TCGTACGGGTGCGGTTCCAGAGAGCGGCGGCTCCGCGAGGCGCGCGCTGGGGGG 62
Qy 208 ProValPheProLysAlaIleGlyLeuHisGlyHisThrGlyIleAlaArgLeuMet 227

```

```
Db      63  GCGGTGTTCCCAAGGCATCGGAGCTGCACGGGCACTCCGGGATCGCGCGCTGGCG 122
Qy      228  LeupheTyAlaArgAlaAlaAlaGlylleValAspProAlaHisGlyThrGlyValAla 247
Db      123  CTGTTCACGGCGCGGGCTGTGGGGCTGTGCACCCGTCGCCCGCGCACGGGGTCGCA 182
Qy      248  AsnAlaGlyLeuValTyPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267
Db      183  AACCGCGGCTCTGCTACTTCAACGGCGCTCTCTGCCCATGTCCGAGGACGACCTGCG 242
Qy      268  TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPhe 287
Db      243  TACGAGTGTGCGCGCGCGGCGGACCTCCGCACCGCTCGGCCCTACGACTTCGGC 302
Qy      288  GlyGlnLeuGlu-----SerThrMetIleAlaHisProLysValAspProGluSerGly 305
Db      303  GGGCAGCTCGCGGGTGGCGGAGCATGATCGGCGACCCCAAGCTTGACCCGGGCTCGGG 362
Qy      306  GluLeuPheAlaLeuSerTyAspValValSerLysProTyLeuLysTyPheArgPhe 325
Db      363  GAGCTGTTCCGGCTCAGTACGAGTGTATCAAGCGGCCCTACCTGAGGTACTTCTAC 422
Qy      326  SerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMet 345
Db      423  CGGCCGAGCGGCGCAAGTCGGCGACGTGGAGATCCCGCTGGAGACCCACCATGGTG 482
Qy      346  HisAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLys 365
Db      483  CACGACTTCGGCATCACGGAGCGGTTCTGTGGTGTGCGCGCACCCAGCGTGTGTTCA 542
Qy      366  LeuProGluMetIleArgGlySerProValValTyAspLysAsnLysValAlaArg 385
Db      543  CTGGCGGAGATGCTCGGGGCGGGTCCCGCTGGTGTGGAGCGCGGCAAGCGTCCGG 602
Qy      386  PheGlyIleLeuAspLysTyAlaGluAspSer-SerAsnIleLysThrIleAspAla 405
Db      603  TTCGGCGTGTGTCGCCAAGTACGCGCGCGAGCGCGCTCGGAGATGGCGTGGTGGAC 662
Qy      405  oAsp 406
Db      663  GGAC 666
```

Search completed: November 16, 2003, 23:12:10
Job time : 3382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:13:32 ; Search time 39 Seconds

(without alignments)
3963.423 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASPTATAVSGRWLGNGHT.....VPYGFHGTFFIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mnc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150	100.0	599	Q9LR7	Q9LR7 arabidopsis
2	3143	99.8	599	Q93ZUS	Q93ZUS arabidopsis
3	2289	72.7	604	Q9M329	Q9M329 solanum tub
4	2280.5	72.4	605	Q24023	Q24023 lycopersico
5	2170.5	68.9	657	Q9M9F5	Q9M9F5 arabidopsis
6	2168.5	68.8	612	Q9FS24	Q9FS24 vigna ungui
7	2165	68.7	615	Q9M6E8	Q9M6E8 phaseolus v
8	2155.5	68.4	601	Q9LP16	Q9LP16 pisum sativ
9	2073.5	65.9	608	Q8LP15	Q8LP15 pisum sativ
10	2066.5	65.6	625	Q9AXZ4	Q9AXZ4 persea amer
11	2044.5	64.9	589	Q9C6Z1	Q9C6Z1 arabidopsis
12	1991	63.2	583	Q49505	Q49505 arabidopsis
13	1977	62.8	569	Q9AXZ3	Q9AXZ3 persea amer
14	1930	61.3	604	Q24592	Q24592 zea mays (m
15	1663.5	52.8	577	Q9LRM7	Q9LRM7 arabidopsis
16	983	31.2	574	Q8LP14	Q8LP14 pisum sativ

ALIGNMENTS

RESULT 1

Q9LR7 PRELIMINARY; PRT; 599 AA.

AC Q9LR7, 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase (Neoxanthin cleavage enzyme).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Col;
RA Iuchi S., Kobayashi M., Shinozaki K.;
RT "Characterization of neoxanthin cleavage enzyme from Arabidopsis
RT thaliana.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028617; BAB01336.1; -
DR EMBL; AB026549; BAB70609.1; -
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.

17 968 30.7 595 10 Q49675
18 959.5 30.5 543 10 Q94IR2
19 945.5 30.0 541 10 Q8LP17
20 938 29.8 538 10 Q9LY63
21 937 29.7 538 10 Q65572
22 922.5 29.3 501 10 Q8GR12
23 837.5 26.6 200 10 Q94EN8
24 798.5 25.3 524 10 Q9AXZ5
25 745.5 23.7 475 16 Q8XV3
26 745 23.7 446 10 Q49895
27 528.5 16.8 483 16 Q9AA32
28 527.5 16.7 456 2 Q9AM11
29 526.5 16.7 456 2 Q93FA4
30 465 14.8 501 16 Q06785
31 464 14.7 503 16 Q9EX45
32 454.5 14.4 502 16 Q05905
33 394 12.5 490 16 P74334
34 382.5 12.1 209 10 Q8VXN9
35 382.5 12.1 251 10 Q8LNK8
36 358 11.4 487 16 Q8DMU6
37 355.5 11.3 497 16 Q8YPB4
38 352 11.2 485 2 Q53353
39 327 10.4 218 10 Q8VXP1
40 322.5 10.2 570 10 Q8VI26
41 319 10.1 520 16 Q8XTA2
42 317 10.1 480 16 P74370
43 311.5 9.9 491 2 Q8RQW2
44 303.5 9.6 490 2 Q52008
45 302.5 9.6 552 10 Q93VD5

Q49675 arabidopsis
Q94IR2 phaseolus v
Q8LP17 pisum sativ
Q9LY63 arabidopsis
Q65572 arabidopsis
Q8GR12 arabidopsis
Q94EN8 lactuca sat
Q9AXZ5 persea amer
Q8XV3 anabaena sp
Q49895 malus domes
Q9AA32 caulobacter
Q9AM11 streptomyces
Q93FA4 streptomyces
Q06785 mycobacteri
Q9EX45 streptomyces
Q05905 mycobacteri
P74334 synechocyst
Q8VXN9 crocus sativ
Q8LNK8 oryza sativ
Q8DMU6 synechococcc
Q8YPB4 anabaena sp
Q53353 pseudomonas
Q8VXP1 crocus sativ
Q8VI26 arabidopsis
Q8XTA2 raietonia s
P74370 synechocyst
Q8RQW2 pseudomonas
Q52008 pseudomonas
Q93VD5 oryza sativ

```
SQ SEQUENCE 599 AA; 65856 MW; 7D513F39945E0CF3 CRC64;
Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 1.6e-235;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAASGRWLGNGHTQPPLSSQSSDLSYCSLPMASRVTRKLNVSALHTPPAL 60
DB 1 MASFTATAAASGRWLGNGHTQPPLSSQSSDLSYCSLPMASRVTRKLNVSALHTPPAL 60
QY 61 HFPKQSSNSPAIVVKPAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLPTADPS 120
DB 61 HFPKQSSNSPAIVVKPAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLPTADPS 120
QY 121 VOIAGNFAPVNEQVARNLPVVGKLPDSIKGVVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
DB 121 VOIAGNFAPVNEQVARNLPVVGKLPDSIKGVVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
QY 181 KFEHGSASACRFTQTNRFVQERQLGRPVFPKAI GELHGHGTIARLMLFYARAAAGIVDP 240
DB 181 KFEHGSASACRFTQTNRFVQERQLGRPVFPKAI GELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVANAGLVYFNGELLAMSEDDLPYQVQITPNGDLKTVGRFDPDGOLESTMIAPKV 300
DB 241 AHGTGVANAGLVYFNGELLAMSEDDLPYQVQITPNGDLKTVGRFDPDGOLESTMIAPKV 300
QY 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
QY 361 QVVPKLPKEMIRGGSPVVDKKNVARGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420
DB 361 QVVPKLPKEMIRGGSPVVDKKNVARGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420
QY 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEDQOVNLEAGM 480
DB 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEDQOVNLEAGM 480
QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTEVKHLYGDNRYGGEPFLPGE 540
DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTEVKHLYGDNRYGGEPFLPGE 540
QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYFGHGTIFGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYFGHGTIFGADDLAKQV 599

RESULT 2
Q93ZU5 PRELIMINARY; PRT; 599 AA.
AC Q93ZU5;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Putative 9-cis-epoxycarotenoid dioxygenase.
GN MOA2.4/AT3G14440.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yanada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk P., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MOA2.4/AT3G14440 (GI:11994214).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY056255; AAL07104.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
DR DIOXYGENASE
SQ SEQUENCE 599 AA; 65813 MW; B9007A2DC1C15506 CRC64;
Query Match 99.8%; Score 3143; DB 10; Length 599;
Best Local Similarity 99.8%; Pred. No. 5.7e-235;
Matches 598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAASGRWLGNGHTQPPLSSQSSDLSYCSLPMASRVTRKLNVSALHTPPAL 60
DB 1 MASFTATAAASGRWLGNGHTQPPLSSQSSDLSYCSLPMASRVTRKLNVSALHTPPAL 60
QY 61 HFPKQSSNSPAIVVKPAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLPTADPS 120
DB 61 HFPKQSSNSPAIVVKPAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLPTADPS 120
QY 121 VOIAGNFAPVNEQVARNLPVVGKLPDSIKGVVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
DB 121 VOIAGNFAPVNEQVARNLPVVGKLPDSIKGVVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
QY 181 KFEHGSASACRFTQTNRFVQERQLGRPVFPKAI GELHGHGTIARLMLFYARAAAGIVDP 240
DB 181 KFEHGSASACRFTQTNRFVQERQLGRPVFPKAI GELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVANAGLVYFNGELLAMSEDDLPYQVQITPNGDLKTVGRFDPDGOLESTMIAPKV 300
DB 241 AHGTGVANAGLVYFNGELLAMSEDDLPYQVQITPNGDLKTVGRFDPDGOLESTMIAPKV 300
QY 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
QY 361 QVVPKLPKEMIRGGSPVVDKKNVARGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420
DB 361 QVVPKLPKEMIRGGSPVVDKKNVARGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420
QY 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEDQOVNLEAGM 480
DB 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEDQOVNLEAGM 480
QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTEVKHLYGDNRYGGEPFLPGE 540
DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTGTEVKHLYGDNRYGGEPFLPGE 540
QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYFGHGTIFGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYFGHGTIFGADDLAKQV 599

RESULT 3
Q9M3Z9 PRELIMINARY; PRT; 604 AA.
AC Q9M3Z9;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Putative 9-cis-epoxycarotenoid dioxygenase.
GN NCED1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RA Burbidge A., Taylor I.B., Thompson A.;
RT "Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ276244; CAB76920.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
```


RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altari H., Bei Q., Chin C., Chou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RA "Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
 RT I.;
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RL Ecker J.R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altari H., Bei B., Chin C., Chou J., Choi E., Conn L.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howng B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC013430; AA71797.1; -;
 DR InterPro; IPR004294; RPE65.
 DR Pfam; PF03055; RPE65; 1.
 SQ SEQUENCE 657 AA; 73015 MW; F41DCEB94806318 CRC64;

Query Match 68.98; Score 2170.5; DB 10; Length 657;
 Best Local Similarity 67.78; Pred. No. 1.7e-159;
 Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

QY 1 MASPTATAVSGRWL-----GNHPTOPPLSSQSDLSYCSSLP-----MASRVT 45
 DB 51 MASPTLLPSTSTQFLDTFTSTSSSRPKLQS-----LSFSTLRNKLWVPCVSSVN 105
 QY 46 KLVNSALHT----PALHFPKROSSPAIVVKPKAKESNTKOMLFOFAAAALDAAE 101
 DB 106 KSSVSSLSQSPPTKPSW---KLCNDVTNLI-PKTNQNP-KLPVQRTAMVLDAYE 160
 QY 102 GFLVSH- KLHLEPKTADPSVQIAGNFAPVNEQVRRLNPVVGKLPDSIKGVYVRNGANP 160
 DB 161 NAMISHERRRHPKTPADPAVQIAGNFFVPEKPVVHNPVTVGPECIQGVYVRNGANP 220
 QY 161 LHEPVTGHHFPDGGMHWAKFEHGSASYACRFTQTNRFVQEROLGRPVFPKALGELHGH 220
 DB 221 LHKPVSGHLLPDGGMHWAVRFDNGSVSYACRFTETNRLVQERECGRPVFPKALGELHGH 280
 QY 221 TGIARLMFLFYARAAAAGIVDPAHGTGVANAGLVYFNGRLAMSDEDLFPYQVQITPNGDLKT 280
 DB 281 LGIAKMLFNTRGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLPHYVKVQTQGDLET 340
 QY 281 VGRFDQQLSTWIAHFKVDPESGELFALSVDVYVKPYLYKYPFSPDGTSPDVEIQLD 340
 DB 341 SGRYDFDQQLASTWIAHFKIDPETRELFAVSVDVYVKPYLYKYPFSTSGEKSPDVEIPLD 400
 QY 341 OPTMHDFAITENFVVVDQVQVFKLPEMIRGSGPVYDKNKVARFGILDKYAESSNIK 400
 DB 401 OPTMHDFAITENFVVDQVQVFKLPEMIRGSGPVYDKNKVARFGILDKYAESSNIK 460
 QY 401 WIDAPDCFCFHLNNAWSEPETDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 460
 DB 461 WIEVPCFCFHLNNAWSEPETDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 520
 QY 461 STRRPIISNEDQVNLNAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 520
 DB 521 STRRPIIS---EQVNLNAGVNRNMLGRKTRAYALALTEPPKVSFGAKVDLSIGIRKY 577
 QY 521 LYGDNRVYGGEPFLFPGEGEEDGYILCFVHDEKTKWSELQIVNAVSVLEATVKLPSRV 580
 DB 578 IYGEKYGGEPLFLPFGEGEEDGGYIMVFVHDEKVKSELQILINAVNMKLEATVTLPSRV 637

QY 581 PYGFHGTGFIAGADDLAKQVV 599
 DB 538 PYGFHGTGFIISKEDLSKQAL 656
 RESULT 6
 ID Q9FS24 PRELIMINARY; PRT; 612 AA.
 AC Q9FS24;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Neoxanthin cleavage enzyme.
 GN CPRD65.
 OS Vigna unguiculata (Cowpea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
 OX NCBI_TaxID=3917;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20317197; PubMed=10859185;
 RA Iuchi S., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K.;
 RT "A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase
 RT involved in abscisic acid biosynthesis under water stress in drought-
 RT tolerant cowpea";
 RL Plant Physiol. 123:553-562 (2000).
 DR EMBL; AB030293; BAB11932.1; -;
 DR InterPro; IPR004294; RPE65.
 DR Pfam; PF03055; RPE65; 1.
 SQ SEQUENCE 612 AA; 67715 MW; E3263A13293B188B CRC64;

Query Match 68.84; Score 2168.5; DB 10; Length 612;
 Best Local Similarity 68.64; Pred. No. 2.1e-159;
 Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

QY 8 AAVSGRWLGNGHTOPPSSQSDLSYCSSLPMAASVTRKLN-VSSALHTPPALHFPKQ- 65
 DB 3 SSASNTWENATLPSPPPKDLPTS-SPTNLLPL--RKTSSNTITCSLOT---LHFPKQY 56
 QY 66 ----SSNSPAIVVKPKAESNT-----KOMLFOFAAAALDAAE 100
 DB 57 QSTSTSTATTTPTPKTTITTTTPPRETNPLSDTNQPLPKWFLQKAAATALLDV 116
 QY 101 EGFVSHSEKHLPLKPTADPSVQIAGNFAPVNEQVRRLNPVVGKLPDSIKGVYVRNGANP 160
 DB 117 ETALVSHERKHLPLKPTADPRVQIAGNFAPVPEHAADQGLPVVGKIPKICIDGVYVRNGANP 176
 QY 161 LHEPVTGHHFPDGGMHWAKFEHGSASYACRFTQTNRFVQEROLGRPVFPKALGELHGH 220
 DB 177 LYEPVAGHHFPDGGMHWAVRFTNGAASACRFTETQLSSEKSLGRPVFPKALGELHGH 236
 QY 221 TGIARLMFLFYARAAAAGIVDPAHGTGVANAGLVYFNGRLAMSDEDLFPYQVQITPNGDLKT 280
 DB 237 SGIAKMLFYARGLFGLVDGSGQMGVANAGLVYFNNHLLAMSEDDLPHYVRITPNGDLTT 296
 QY 281 VGRFDQQLSTWIAHFKVDPESGELFALSVDVYVKPYLYKYPFSPDGTSPDVEIQLD 340
 DB 297 VGRYDFNGQLNSTWIAHFKLPDVGDLHALSYDVYQKPYLYKYPFSPDGTSPDVEIPLK 356
 QY 341 OPTMHDFAITENFVVVDQVQVFKLPEMIRGSGPVYDKNKVARFGILDKYAESSNIK 400
 DB 357 EPTMHDFAITENFVVVDQVQVFKLTEMITGSGPVYDKNKTSRFGILHKNKADANAR 416
 QY 401 WIDAPDCFCFHLNNAWSEPETDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 460
 DB 417 WIDAPDCFCFHLNNAWSEPETDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 476
 QY 461 STRRPIISNEDQVNLNAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 520
 DB 477 STRRPIIS---DAQVNLNAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLSIGEVKY 535
 QY 521 LYGDNRVYGGEPFLFPGEGEEDGYILCFVHDEKTKWSELQIVNAVSVLEATVKLPSRV 580

```
Db 536 MYGEKFGGEPLFLP-NGQDDGGYILAFVHDEKEWKSELOIVNAQNKLKLEASIKLPSRV 594
QY 581 PYGPHGTFTICADDLAKQ 597
Db 595 PYGPHGTFTIHSKDLRKQ 611

RESULT 7
Q9M6E8
ID Q9M6E8 PRELIMINARY; PRT; 615 AA.
AC Q9M6E8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Top Crop;
RX MEDLINE=20079657; PubMed=10611388;
RA Qin X., Zeavaart J.A.;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
step of abscisic acid biosynthesis in water-stressed bean.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:15354-15361 (1999).
DR EMBL; AF190462; AAF26356.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 615 AA; 68075 MW; 0CC10F862D7DE130 CRC64;

Query Match 68.7%; Score 2165; DB 10; Length 615;
Best Local Similarity 69.2%; Pred. No. 4e-159;
Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

QY 11 SGRWLGNGHTQPLSSQSDLSYSS-----LPMASVTRKLN-VSSALHTPPALHPFK 64
Db 6 SNTWI--NTTLSSSSPKDLASTSSPTTLLPFKRSSTNTNITCSLQI--LHYPK 60
QY 65 Q-----SNSPAIVVYPKAKE-----SNTKQ-----NMLFORAAALDAAE 101
Db 61 QYQPTSTTTTPTIKPTTTTTTTPHRETRKPLSDTKQFPQKWNFLQKAAATGLDMVE 120
QY 102 GFVVSHEKHLPLPKTADPSVQIAGNFPVNEQVRNLPVVGKLPDSIKGVYVRNGANPL 161
Db 121 TALVSHESKHLPLPKTADPKVQIAGNFPVPEHAADQALPVVGKI PKCIDGVYVRNGANPL 180
QY 162 HEPVTGHHFFDGMVHAYKFEHGSASYACRTQTNRFRVQEROLGRPVFPKAI GELHGT 221
Db 181 YEPVAGHHFFDGMVHAYKFTNGAASYACRTETQRLAQEKSLGRPVFPKAI GELHGS 240
QY 222 GIARLMLFARAAGVDPHAGTGVANAGLVFENGELAWSDDL PYQVQITPENGDLKT 281
Db 241 GIARLLFTVARSFQVLDGSHGVANAGLVFNHLLAWSDDL PYHVRITNSGDLTV 300
QY 282 GREDFDQGLESTMIAPKVPDPSGELFALSVDVSKPYLKYFRFSPDGKSPDVEIQLDQ 341
Db 301 GRYDFNGQLNSTMIAPKLPDVPNGDLFALSVDVQKPYLKYFRFSGADGVKSPDVEIPLKE 360
QY 342 PTMHDFAITENFVVPDQVFPKLPDMIRGSPVYVDKNKVARGLDKYAEADSNIKW 401
Db 361 PTMHDFAITENFVVPDQVFPKLPDMIRGSPVYVDKNKTSRFGILDKNADKANMRW 420
QY 402 IDAPDFCFHLMNAWEPEPTEDEVVWIGSMTTPDSIFNESDENLKSLSVSEIRLNLTGGS 461
Db 421 IDAPECFCHLMNAWEPEPTEDEVVWIGSMTTPADSIFNECDLSLSVSEIRLNLTGKS 480
QY 462 TRRPIISNEDQVNLBAGMVRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTFGEVKKHL 521
```

```
Db 481 TRRPIIS-DABQVNLBAGMVRNMLGRKTKFAYLALAEWPVKVSGFAKVDLTFGEVQKTM 539
QY 522 YGDNRYGGEPLFLPGEGBEDEGYILCFVHDEKTKWSELQIVNAVSELYEATVKLPSRVP 581
Db 540 YGEEKGGEPLFLP-NGEEDGGYILAFVHDEKEWKSELOIVNAQNKLKLEASIKLPSRVP 598
QY 582 YGPHGTFTICADDLAKQ 597
Db 599 YGPHGTFTIHSKDLRKQ 614

RESULT 8
Q8LP16
ID Q8LP16 PRELIMINARY; PRT; 601 AA.
AC Q8LP16;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Nine-cis-epoxycarotenoid dioxygenase2.
GN NCED2.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Axillary bud;
RA Nakako A., Mori H.;
RT "Isolation and characterization of a cDNA encoding a nine-cis-
epoxycarotenoid dioxygenase2 (PNCED2) from Pisum sativum.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080192; BAC10550.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 601 AA; 67287 MW; F4EDC257867CF195 CRC64;

Query Match 68.4%; Score 2155.5; DB 10; Length 601;
Best Local Similarity 68.7%; Pred. No. 2.1e-158;
Matches 422; Conservative 61; Mismatches 100; Indels 31; Gaps 8;

QY 1 MASFTATAVSGRWLGNHTQPLSSQSDLSYSSQSDLSYSSLPASVTRKLNVSALHTPPAL 60
Db 1 MATFTASP--SNTW-----INTKSSRTSSSILLNKKRSTNTITCSLQI--TL 47
QY 61 HFPK-----QSSNSPAIVVYPKAKE-----ESNTKQ-----NLFORAAALDAAEGL 104
Db 48 PFPKYQPSQSTNTTTTLLPTRETKNLPSTNPKLHKQEKQKLLQAAAATLDFVETTL 107
QY 105 VSHEKHLPLPKTADPSVQIAGNFPVNEQVRNLPVVGKLPDSIKGVYVRNGANPLHEP 164
Db 108 IKOESKHLPLPKTADPSVQIAGNFPVPEHPTQNLPTITGKLPKIDGVYLRNGANPLHEP 167
QY 165 VTGHHFFDGMVHAYKFEHGSASYACRTQTNRFRVQEROLGRPVFPKAI GELHGTGIA 224
Db 168 VAGHHFFDGMVHAYKFTNGSVYSORFTETHLAQEKALGRPVFPKAI GELHGSIA 227
QY 225 RLMLFYARAAGVDPHAGTGVANAGLVFENGELAWSDDL PYQVQITPENGDLKT VGRF 284
Db 228 RLMLFYARSLCGLVDGTHGMGVANAGLVFENNLAWSDDL PYHVRVTENGDLTTCRY 287
QY 285 DFQDQGLESTMIAPKVPDPSGELFALSVDVSKPYLKYFRFSPDGKSPDVEIQLODPTM 344
Db 288 DFNDQLKSTMIAPKVPDVKNIYALSYDVQKPYLKYFRFSDNGVSKSPDVEIPLAEP 347
QY 345 MHDFAITENFVVPDQVFPKLPDMIRGSPVYVDKNKVARFGILDKYAEADSNIKWIDA 404
Db 348 MHDFAITENFVVPDQVFPKLPDMIRGSPVYVDKESKVSFRFGVLSKNAENASEMKWIDA 407
QY 405 PDCEFCFCHLMNAWEPEPTEDEVVWIGSMTTPDSIFNESDENLKSLSVSEIRLNLTGSTR 464
```

```

Db 408 PECFCFLMNAWEEPEDEVVIGSWTTPADSIIFNECDSELSKSVLSVLEIRLNLTGKSTR 467
QY 465 PITSNEDQVNLQAGMVRNMLGRKTKFAYLALAEPPKVSQFAKVDLTGTVKKHLYCD 524
Db 468 AII-QSEHNLQAGMVRNMLGRKTKFAYLALAEPPKVSQFAKVDLTGTVKKHLYGE 526
QY 525 NRYGGPFLPLPGGGE-EDBEGYILCFVHDEKTKWSELOQVNAVSLVEATVKLPGRVPYG 583
Db 527 NREGPFLPNEDSENEDDGYILTTFVHDEKWKSEIQVNAVNLKLEACIPLPSRVPYG 586
QY 584 FHGTFCADDLAKQ 597
Db 587 FHGTFTHSKELEQ 600

RESULT 9
Q8LP15 PRELIMINARY; PRT; 608 AA.
ID Q8LP15
AC Q8LP15
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nine-cis-epoxycarotenoid dioxygenase3.
GN NCED3
OS Pism. sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pism.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Axillary bud;
RA Nakako A., Mori H.;
RT "Isolation and characterization of a cDNA encoding a nine-cis-
RT epoxycarotenoid dioxygenase3 (PNCED3) from Pism. sativum."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080193; BAC10551.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 608 AA; 67978 MW; 73E80913AAA26C4D CRC64;

Query Match 65.9%; Score 2075.5; DB 10; Length 608;
Best Local Similarity 68.3%; Pred. No. 3.4e-152;
Matches 403; Conservative 64; Mismatches 98; Indels 25; Gaps 5;

QY 24 LSSQSSDLSCSLPVAASVTRKLVNSSLHTTPALHFPKQSSNSP-----AIVV 74
Db 28 ISSSSSTSFNDKSTLKLKKPKRKLLOCAVHSFSLDYPKQSYKQPLINKENDTETIIH 87
QY 75 KPKAESNTKQNLQFQAAAAALDAAGFLVSHKHLPLPKTADPSVQIAGNFAPVNEQP 134
Db 88 KPK----ESSQWNLQKAAALANWFESALLSRELOVPLPKTSDPRIQIAGNFAPVPEQP 143
QY 135 VRNLPLVVKLPDSIKGVYVNGANPLHEPTVGHFFDGGDMHVAKFEGCSASYACRFT 194
Db 144 VVHSLFVTKGKPRCVNGYVYVNGANPMPFPPVSGHLLFDGGDMHVAFTNDGVASYACRFT 203
QY 195 QTNREVRQRLGRPVFPKAIAGELHGHGTGIARLMLFYARAAAGIYDPAHGTGVANAGLVYF 254
Db 204 ETERLVQERELGRAMFPKAIAGELHGHGTGIARLMLFYARSLCGIYDHRGGGVANAGLVYF 263
QY 255 NGRLLAMSEDDLPYQVQITPNGDLTKVGRFPDQGLESTMTAHKVPDPESGELFALSVDV 314
Db 264 DGKLLAMSEDDLPYELQTPYGDLTGVRYSFLDQLHSTMTAHKPIDIPISGELFALSVE- 322
QY 315 VSKPYLVFRSPDGTGKPDVEIQDQPTMHDFAITENFVVPDQVVKLPKPMIRGGG 374
Db 323 VARPLVKVFRSPDGKSPDVEIRLSVFTMIHDAITENYVYIPDHQVVKLEEMIKGCS 382
QY 375 PVYDKNKVARPGLDKVAEASSNTKIWDADCFCHLWNAWEPETDEVVVIGSCMTTPR 434
Db 383 PVIFDGAKSRGVLGPKYAKDASSIIWVDSPTDFCHLWNAWEEPEADEIIVIGSCMTTP 442

```

```

QY 435 DSIFNESDENLKSVLSEIRLNLTGSTRPPIISNEDQVNLQAGMVRNMLGRKTKFAY 494
Db 443 DSIFNESDENLKSVLSEIRLNLTGSTRPPIISNEDQVNLQAGMVRNMLGRKTKFAY 498
QY 495 LALAEPPKVSQFAKVDLTGTVKKHLYGNRYGGPFLPLPG-----GGDESGYIL 547
Db 499 LAVAEPPKVSQFAKVDLESGEVKKHSGYDGRFGGPFPLPFRSGRCEYENEDGGYIM 558
QY 548 CFVHDEKTKWSELOQVNAVSLVEATVKLPGRVPYGFHGTFCADDLAKQ 597
Db 559 ALVHDERCKSELOQVNAVSLVEATVKLPGRVPYGFHGTFCADDLAKQ 608

RESULT 10
Q9AXZ4 PRELIMINARY; PRT; 625 AA.
ID Q9AXZ4
AC Q9AXZ4
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED3
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Lula;
RA Chernys J., Zeevaert J.A.D.;
RT "Abasic acid in avocado fruit."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF224671; AAK00623.1;
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF03055; RPE65; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Dioxygenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F33042D5F2CB6 CRC64;

Query Match 65.6%; Score 2066.5; DB 10; Length 625;
Best Local Similarity 64.9%; Pred. No. 1.8e-151;
Matches 392; Conservative 74; Mismatches 93; Indels 45; Gaps 6;

QY 38 LPMASRVTNRKLVNSSLHTTP-----ALHFP-----KQSS 67
Db 22 LPISKNLSRPKNFIMLKHTPLIQCSHSPSSSSAAVLHLPKQPKTSKPSIKKGEKSS 81
QY 68 NSPAIVVVKPKAKESNTKQ-----NVLQFQAAAAALDAAGFLVSH--EKLHPLPKT 116
Db 82 LTPSIEKPNPGSHQVKTQSGPNRVGNPNWIFQRTAAAFALDAIEKLIARVLERHPLPKT 141
QY 117 ADPSVQIAGNFAPVNEQPVRRNLPVVKLPDSIKGVYVNGANPLHEPTVGHFFDGGDM 176
Db 142 ADPEVQIAGNFAPVAEHPVQHGIPVAGRIPRCLDGVYVNGANPLFEBIAGHFFDGGDM 201
QY 177 VNAVFEHGSASYACRFTQTRFQEROLGRPVFPKAIAGELHGHGTGIARLMLFYARAAAG 236
Db 202 IHAVFRNGSASYSRYTETRLVQEROLSRPIPKAIAGELHGHGTGIARLMLFYTRGLFG 261
QY 237 IYDPAHGTGVANAGLVYVNGRLLAMSEDDLPYQVQITPNGDLTKVGRFPDQGLESTMTA 296
Db 262 LVNADEGKVANAGLVYVNRLLAMSEDDLPVHVEITPSGDLTKVGRHDFDQLRSSMIA 321
QY 297 HPKVPDPESGELFALSVDVVKSPYLYKFRSPDGTGKSPDVEIQDQPTMHDFAITENFV 356
Db 322 HPKLDPESGELFALSVDVARKPYLYKFRFAPDGWKSVDVEIPDRFTMIHDAITENFV 381
QY 357 VPDQVVKLPKPMIRGGSPVYVKNKVARFGITDKVAEDSSNKKIWDADCFCHLWNAW 416
Db 382 IPDQVVKLEEMIRGGSPVYDKNKTSRFGILLPKIAPDASEMIWVADPDCFCHLWNAW 441

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:12:47 ; Search time 17 Seconds
(without alignments)
1657.000 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAAAGSRMLGNHT.....VPYGFHGTTCGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	3.8	616	1 NXF1_COTJA	P58797 coturnix co
2	114	3.6	870	1 DCSO_YEAST	P25623 coturnix co
3	113	3.6	466	1 DNAA_PROMI	P22837 proteus mir
4	107.5	3.4	658	1 PAK1_SCHPO	P50527 schizosacch
5	106	3.4	414	1 STG6_HUMAN	O94864 homo sapien
6	106	3.4	3562	1 PGCV_CHICK	Q90953 gallus gall
7	105	3.3	783	1 CAP5_MOUSE	P55284 mus musculu
8	105	3.3	2504	1 PAS_HUMAN	P43127 homo sapien
9	104.5	3.3	1079	1 APCE_PREDI	P16566 fremyella d
10	103	3.3	775	1 SYFB_AQUAE	O67620 aquifex aeo
11	102	3.2	513	1 M113_MOUSE	Q8Bgt6 mus musculu
12	102	3.2	2493	1 CYAA_USTWA	P49606 ustilago ma
13	101.5	3.2	520	1 GLGS_BRANA	Q9m462 brassica na
14	101.5	3.2	562	1 SVR_XANCP	O8p455 xanthomonas
15	101.5	3.2	565	1 SCH1_SCHPO	O14335 schizosacch
16	101	3.2	512	1 GLGT_VICFA	P52417 vicia faba
17	100	3.2	3770	1 ACVS_EMENI	F27742 emeritella
18	99.5	3.2	383	1 PHYC_BACSU	O31097 bacillus su
19	99.5	3.2	1256	1 ATL_STRAU	P52081 staphylococ
20	99.5	3.2	4705	1 FAT2_DROME	Q9vw71 drosophila
21	98	3.1	741	1 MAS2_MYCTU	O50596 mycobacteri
22	98	3.1	1633	1 VP74_CAEEL	Q09321 caenorhabdi
23	97.5	3.1	518	1 CP74_ARATH	Q98242 arabidopsis
24	97.5	3.1	691	1 VP33_YEAST	P20795 saccharomyc
25	97.5	3.1	730	1 EF2_METAC	Q8trc3 methanosarc
26	97	3.1	385	1 OXDC_BACSU	O34714 bacillus su
27	97	3.1	636	1 KDGP_SCHPO	O14019 schizosacch
28	96.5	3.1	426	1 UCR2_SCHPO	P78761 schizosacch
29	96.5	3.1	503	1 DLTA_BACSU	P39581 bacillus su
30	96.5	3.1	862	1 SMF2_YEAST	P32567 saccharomyc
31	96.5	3.1	3587	1 TYCE_BREPA	O30408 b tyrocidin
32	96	3.0	462	1 DNAA_YERPE	Q829u7 yersinia.be
33	95.5	3.0	865	1 NRFA_PENUR	Q92269 penicillium

ALIGNMENTS

RESULT 1

ID	NXF1_COTJA	STANDARD;	PRT;	616 AA.
AC	P58797			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).			
DE	NXF1			
GN	NXF1			
OS	Coturnix coturnix japonica (Japanese quail).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Coturnix			
OX	NCBI_TaxID=93934;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND MUTAGENESIS OF GLU-246.			
RX	MEDLINE=21257892; PubMed=11358864;			
RA	Coturn G.A., Wiegand H.L., Kang Y., Ho D.N., Georgiadis M.M.,			
RA	Cullen B.R.;			
RT	Using viral species specificity to define a critical protein/RNA interaction surface."			
RL	Genes Dev. 15:1194-1205 (2001).			
CC	-1- FUNCTION: Involved in the nuclear export of mRNA from the nucleus to the cytoplasm.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			
CC	-1- DOMAIN: The leucine-rich repeats and the NTF2-domain are essential for the export of mRNA from the nucleus (By similarity).			
CC	-1- MISCELLANEOUS: Changing a single residue, Glu-246 to Arg, the residue found in human NXF1, fully rescues both constitutive transport element (CTE) function and CTE binding.			
CC	-1- SIMILARITY: BELONGS TO THE NXF FAMILY.			
CC	-1- SIMILARITY: Contains 4 leucine-rich (LRR) repeats.			
CC	-1- SIMILARITY: Contains 1 NTF2 domain.			
CC	-1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.			
CC	-1- SIMILARITY: Contains 1 UBA-like domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF343749; AA58910.1; -			
DR	InterPro; IPR001611; LRR.			
DR	InterPro; IPR003603; LRRcap.			
DR	InterPro; IPR002075; NTF2.			
DR	InterPro; IPR005637; TAP_C.			
DR	Pfam; PF00560; LRR; 2.			
DR	Pfam; PF021136; NTF2; 1.			
DR	Pfam; PF03943; TAP_C; 1.			
DR	SMART; SM00446; LRRcap; 1.			
DR	PROSITE; PS50177; NTF2_DOMAIN; 1.			
KW	Transport; mRNA transport; Nuclear protein; Repeat;			

[illegible]

GN PAK1 OR SHK1 OR ORB2 OR SPBC1604.14C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
[1]
RN
RN
RN SEQUENCE FROM N.A.
RP MEDLINE=96112805; PubMed=8846783;
RX Ottilie S., Miller P.J., Johnson D.I., Creasy C.L., Sells M.A.,
RA Bagrodia S., Forsburg S.L., Chernoff J.;
RT "Fission yeast pak1+ encodes a protein kinase that interacts with
RT cdc42p and is involved in the control of cell polarity and mating.";
RL EMBO J. 14:5908-5919(1995).
[2]
RN
RN SEQUENCE FROM N.A.
RP Marcus S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Pest N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gallies S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,
RA Heiroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward I., Volktaert G., Kert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mostl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux C., Leclerc V., Mortier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Dona R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Snipakovits G., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
Nature 415:871-880(2002).
[4]
RN
RN SEQUENCE OF 119-658 FROM N.A.
RX MEDLINE=95320235; PubMed=759708;
RA Marcus S., Polverino A., Chang E., Robbins D., Cobb M.H.,
RA Wigler M.;
RT "Shk1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian
RT p56PAK protein kinases, is a component of a Ras/Cdc42 signaling
RT module in the fission yeast Schizosaccharomyces pombe.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6180-6184(1995).
CC -!- FUNCTION: FORMS AN ACTIVATED COMPLEX WITH GTP-BOUND RAS-LIKE
CC CDC42, ACTS IN SIGNAL TRANSDUCTION. INVOLVED IN THE CONTROL OF
CC CELL POLARITY AND MATING. MAY INTERACT WITH BYR2.
CC -!- PTM: AUTOPHOSPHORYLATED ON SERINE RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE SRP/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -!- SIMILARITY: Contains 1 CRIB domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

DR EMBL; U22371; AAC49125.1; -.
DR EMBL; AL034433; CAA22347.1; -.
DR EMBL; L41552; AAB52609.1; -.
DR FJ; S60170; S60170.
DR FJ; T39500; T39500.
DR HSP; Q00534; IBI8.
DR GeneDB; Spombe; SPSC1604.14c; -.
DR InterPro; IPR000095; PAKBox/RhoBndng.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00786; PDB; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR P-DB; PD00001; Prot kinase; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS00108; CRIB; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 147 160 CRIB.
FT DOMAIN 85 88 POLY-SER.
FT DOMAIN 234 237 POLY-SER.
FT DOMAIN 246 249 POLY-SER.
FT DOMAIN 263 267 POLY-SER.
FT DOMAIN 386 637 PROTEIN_KINASE.
FT NP_BIND 392 400 ATP (BY SIMILARITY).
FT BINDING 415 415 ATP (BY SIMILARITY).
FT ACT_SITE 505 505 BY SIMILARITY.
FT CONFLICT 492 495 GLOH -> LVSD (IN REF. 1).
FT CONFLICT 537 537 R -> P (IN REF. 1).
FT CONFLICT 537 537 R -> P (IN REF. 1).
SQ SEQUENCE 658 AA; 73258 MW; 69D72E5C25021E5 CRC64;

Query Match
Best Local Similarity 3.4%; Score 107.5; DB 1; Length 658;
Matches 88; Conservative 61; Mismatches 157; Indels 121; Gaps 20;

Qy 18 NHTQPLSSQSSDL-----SYCSSLPASRVTRKLVSSALHTPPALHFPKQS 67
Db 237 SHLOPTATSSSLYPSRPAPTPASSSSPLSSQTVKTTTNSASRQSPVLVSKSTD 296

Qy 68 N-----SPAIVVFKAKESNTKQ-----NLFORAAAAALDAAEGFLVSHKHLPLPKTA 117
Db 297 NIIRSHSPVLLTPQTLSTSEKH:RPNNSTPYQRRATSTK-----PKAV 341

Qy 118 DPSVQIAGNPAFVNEQVRRNLVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMV 177
Db 342 ATPQKEAPSAPRLQKRAPRO-----QSNDSAVLAKLOSICNPKNPTLLYRNF----- 389

Qy 178 HAVKFEHGSAS--YACRFTQNRVQERQLGRPVFPK---AIGEL-----HGHGTGIARLM 227
Db 390 --VKGQSGDVYSARQVGNLSVAIKMNIQPKKEFVNEILLVMKSHHKNIVNFI 447

Qy 228 --LYARAAGIVDPAHGTG-----VANAGLVYFNGRLLMSGDDLPVQVITPNG-----D 277
Db 448 DTFYKSELMMVMYMEGGSLEVTYVNTNL--SEGQIAACKETLEGQLHENGIVHRD 505

Qy 278 LKT-----VGRFDGDOLES-----TWIAHP-----KVDPEP 304
Db 506 IKSDNILLSLQDGIKLTDFGCAQIDSNMKTETTMVGTFFYMAPEVVTREYGFKVDVMS 565

Qy 305 GELFALSYDVVSKPYLK-----YFRFSDGTQKSPDVEIQDQTM-----HDAITENF 354
Db 566 LGIMAIEVVEGPPYLNENPLRALYLIATIGTP-----KISRPELLSSVFFHDP-LGSKL 618

Qy 355 VVVPDQ 361
Db 619 TVNPKQR 625

```

```

RESULT 5
ST6G_HUMAN STANDARD; PRT; 414 AA.
AC O94864; O94864;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE STAGA complex 65 Gamma subunit (Adenocarcinoma antigen ART1).
GN KIAA0764.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=20441578; PubMed=10987294;
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes
infiltrating into a lung adenocarcinoma.";
RL Cancer Res. 60:4830-4837(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [3]
RP SEQUENCE OF 6-19, AND IDENTIFICATION AS PART OF THE STAGA COMPLEX.
RX MEDLINE=21448975; PubMed=11564863;
RA Kundu T.K., Chait B.T., Tjernberg A., Lyman E.S., Gamper A.M.,
RA Martnez E., Palhan V.B., Tjernberg R.G.;
RT "Human STAGA complex is a chromatin-acetylating transcription
coactivator that interacts with pre-mRNA splicing and DNA
damage-binding factors in vivo.";
RL Mol. Cell. Biol. 21:6782-6795(2001).
CC -! SUBUNIT: Component of the STAGA transcription coactivator-HAT
complex.
CC -! SURCELLULAR LOCATION: Nuclear.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=O94864-1; Sequence=Displayed;
CC IsoId=O94864-2; Sequence=VSP_003974;
CC Note=No experimental confirmation available;
CC -! TISSUE SPECIFICITY: Expressed at high levels in adenocarcinomas
and gliomas and low in esophageal cancers and malignant
hematological disease. Also expressed at high level in the thymus,
low in peripheral blood mononuclear cells, and lowest in the
stomach, small intestine, and skeletal muscle.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; AF197954; AAG28523.1; -.
CC EMBL; AF224759; AAG47636.1; -.
CC EMBL; AB018307; BAA34484.1; -.
CC GO; GO:0005634; C:nucleus; NAS.
CC InterPro; IPR006565; BTP.
CC SMART; SM00576; BTP; 1.
CC Transcription regulation; Nuclear protein; Antigen;
CC Alternative splicing.
CC VARSPLIC 1 4 MNLQ -> ML (in isoform 2).
CC FTId=VSP_003974.

```


FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 67 69 KDO -> GKIK (IN REF. 2).
 SQ SEQUENCE 783 AA; 87847 MW; 240AF2D663BCE71C CRC64;

Query Match
 Best Local Similarity 17.7%; Score 105; DB 1; Length 783;
 Matches 128; Conservative 101; Mismatches 242; Indels 254; Gaps 32;

QY 60 LHFPKQSNPAIVVVKPAKSNYQKMLFQRAAAA-----ALDAAGFLVSEKHLHP----- 112
 DB 52 MHIDEKQESLPHYKYDQSNVNRQNAKYVLOGEFAGKIFGVDANTGNVLAERLDREKVS 111
 QY 113 -----LPKADPSVQ-----IAGNPAPVNEQFVRRNLPPVGVKLPDSIKGVY 153
 DB 112 EYFELTALIVDKTNKNLBPQSFVTKVHDINDNMFVSHQVFNASVPMSAIGTSVIRVT 171
 QY 154 VRNGANP-----LHEPVTQHFF--DGDGNVHAVKFERGSASYACRFQTQVRFVUER 203
 DB 172 AYVADDDPTVAGHATVLYQIVKNEYFSIDNSGLI-----FTKIKNLDREK 216
 QY 204 QLGRPVFEPK--AIGELHGHGTGIALMLFYARAAAAGIVDPAHGTGVANAGLVYFNGRLLA 260
 DB 217 QAEYKIVVETQALG-LRGESATVMI-----RL-- 245
 QY 261 MSEDLLPYQVQITPNDGLTKVGRFDFGQLESTIAHP-----KYDPSGELFALSVDV 315
 DB 246 --ED-----INDNFPVFTQYTFSPV-EDIRVGKPLGFLTVDPDPQNRMTKYSIM 295
 QY 316 SKPYLKYFRFSP-----GTSPDVE-----IQLDQPTMHDF----- 348
 DB 296 QEYRDTFTIETDPKNEGIIKPTSLDYEVIQOYTFYIEATDPIRVEYLSSTSGKKA 355
 QY 349 AITENFVVVPDQV-----FKLPEMIRG--GSPVVYDKNKVAR-FGILDKYAEDSS- 397
 DB 356 MVTINVLVDPEPVPQRFHFYFKLPENQKKPLIGTVAKDPDKAQRSGYSIRKTSRQ 415
 QY 398 -----NIKMTADAPDCFFHLNAAWEPETDE-----VVVIGSOWT 432
 DB 416 FFRITKQNIYNEKELDRITYAWYLN--TVANELDSRNPVKGSIVGVYIEVDENDN 473
 QY 433 PPDISI-----FNESDENLKSVLSEIRLMKTGSETRPIISNEQQ 473
 DB 474 PFBAQYEPKVCENAAQGLVQISATDKDVVNPKEFKALK-NEDSNFTLINHNDT 532
 QY 474 VNL--BAGMVNRNMLGRKTKPAYLALA---EPMKVGSGFAKVDLTGTEVKKHLYGDNRYG 528
 DB 533 ANITVKYGFNR---EHAKEHYLPVLISDNGVPSLTGTS--TLTVGVCKNEQGEFTFC 586
 QY 529 GEPLFLPGEGBEGEYILCFV-----HDE 553
 DB 587 EEMAAQAGVSIQALVAIFLCILITVITLLIILRRIRKQAHAKSKSALEIHEQLVYDE 646
 QY 554 K-----TWKSELOIVNAV-----SLEVEATVKLPSPVPYGFHGTGIGADD 593
 DB 647 EGGGEMDTTSDYDVSVLNSVRAGGSKPLRSTMDARPAYTQVQKPPPLAPFLGHG--GPRE 703
 QY 594 LAKOV 598
 DB 704 MATMI 708

RESULT 8
 FAS_HUMAN
 ID_FAS_HUMAN STANDARD; PRT; 2504 AA.
 AC P9327;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.41; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.1.14]
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.1.14
 GN FASN OR FAS.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96004605; PubMed=7567999;
 RA Jayakumar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,
 RA Abu-Elheiga L., Chirala S.S., Wakil S.J.,
 RT "Human fatty acid synthase: properties and molecular cloning";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).
 RN [2].
 RP SEQUENCE OF 753-758 AND 1285-1297.
 RX MEDLINE=94294385; PubMed=8022791;
 RA Khatjda F.P., Jenner K., Wood F.D., Hennigar R.A., Jacobs L.B.,
 RA Dick J.D., Pasternack G.R.;
 RT "Fatty acid synthesis: a potential selective target for
 RT antineoplastic therapy";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).
 CC -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
 CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
 CC ACYL CARRIER PROTEIN.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
 CC acetyl-[acyl-carrier protein].
 CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
 CC malonyl-[acyl-carrier protein].
 CC -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
 CC [acyl-carrier protein].
 CC -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
 CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
 CC -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
 CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
 CC -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
 CC carrier protein] + oleate.
 CC -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
 CC -!- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND
 CC LIVER.
 CC -!- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY
 CC MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTHETHEINE CONTENT OF
 CC THE PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U26644; AAC50259.1; -;
 CC PIR; A57788; A57788.
 CC Genew; HGNC:3594; FASN.
 CC MIM; 600212; -;
 CC InterPro; IPR001227; Ac transferase.
 CC InterPro; IPR002085; Adh_zn family.
 CC InterPro; IPR007994; Ketoacyl-synt.
 CC InterPro; IPR006163; PP_bind.
 CC InterPro; IPR006162; Ppantrn attach.
 CC InterPro; IPR000051; SAM_bind.
 CC InterPro; IPR001031; Thioesterase.
 CC Pfam; PF00698; Acyl_transf; 1.
 CC Pfam; PF00107; Adh_zinc_N; 1.
 CC Pfam; PF00109; ketoacyl-synt; 1.
 CC Pfam; PF02801; ketoacyl-synt_C; 1.
 CC Pfam; PF00550; pp-binding; 1.
 CC Pfam; PF00975; Thioesterase; 1.

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sardinia A., Schneider C., Sempé C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan A., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002);
CC -!- SUBUNIT: Binds to Rab3 (By similarity).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK07217; BAC36689.1; -;
CC EMBL; AK076763; BAC36472.1; -;
CC FT NON TER 1 75 POLY-GLU.
CC FT DOMAIN 67 75 POLY-GLU.
CC FT DOMAIN 459 462 POLY-GLU.
CC SQ SEQUENCE 513 AA; 55721 MW; 70CD36AC49CB061 CRC64;
Query Match 3.2%; Score 102; DB 1; Length 513;
Best Local Similarity 22.8%; Pred. No. 5.3;
Matches 66; Conservative 35; Mismatches 82; Indels 106; Gaps 17;
QY 18 NHTQPLSSSSDLSYSSLPMSRVTRKLNVSALHT-----PPALHPKQSSNP 70
DB 149 SHSEPP-----SATPSALSVESLSSESSHTANAEPLPAV--PK-SSSDP 193
QY 71 AI-----VVKPKAESNTKQMLFORAAALDA-AEG----- 102
DB 194 AVHVPGTGTSNSVTPSANGSLSSGELQPSGEQMLQARTKGSAGTSTKPFSGATPT 253
QY 103 FLVSHKHLPLK--TADPSVQIAGNFA--PVNEQPVRLNLPVVGKLPDSIKGVYVRNG 157
DB 254 PELLAGDR-NPAPPVGSASPOLQTKSCKENPFNRKPSASPTVRK-----ATKG 303
QY 158 ANPLHEPVTGHHFFDGDGQVHVAKPEHGSASYACFTQTNRVQERQLGRVPFPAIGEL 217
DB 304 AKVPAPPGHGF-----PLIK-----RKVQADQYIPEEDI-----XGEM 338
QY 218 HGHGTIARLMFLYARAAGIVDPANGTGVANAGLVFNGLRLANSEDDL 266
DB 339 ---DNIERQL--DALEHSGVLEELKRGANEG-----SEDDM 371
RESULT 12
CYAA_USTVA STANDARD; PRT; 2493 AA.
AC P49606;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
DE cyclase).
GN UAC1 OR REM1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=518;
EX MEDLINE=95087882; PubMed=7995519;
RA Gold S., Duncan G., Barrett K., Kronstad J.W.;
RT "cAMP regulates morphogenesis in the fungal pathogen Ustilago
RT maydis";
RL Genes Dev. 8:2895-2816 (1994).
CC -!- FUNCTION: Plays essential roles in regulation of cellular
CC metabolism by catalyzing the synthesis of a second messenger,
CC cAMP.
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
CC -!- SIMILARITY: Contains 21 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 PP2C-like domain.
CC -!- SIMILARITY: Contains 1 Ras-associating domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L33918; AA557469.1; -;
CC PIR; A55481; A55481. G cyclase.
CC InterPro; IPR001854; G cyclase.
CC InterPro; IPR001511; LRR.
CC InterPro; IPR003591; LRR typ.
CC InterPro; IPR001932; PP2C-like.
CC InterPro; IPR000159; RA domain.
CC Pfam; PF00211; Guanylate_cyc; 1.
CC Pfam; PF00560; LRR; 16.
CC Pfam; PF00481; PP2C; 1.
CC SMART; SM00044; CYCC; 1.
CC SMART; SM00369; LRR_TYP; 2.
CC SMART; SM00332; PP2C; 1.
CC SMART; SM00314; RA; 1.
CC PROSITE; PS01225; GUANYLATE_CYCLASES_2; 1.
CC PROSITE; PS02000; RA; 1.
CC Lyase; Repeat; Leucine-rich repeat; cAMP biosynthesis; Metal-binding;
KW Magnesium.
KW DOMAIN 970 1072 RAS-ASSOCIATING.
FT REPEAT 1084 1106 LRR 1.
FT REPEAT 1108 1131 LRR 2.
FT REPEAT 1132 1155 LRR 3.
FT REPEAT 1157 1178 LRR 4.
FT REPEAT 1179 1201 LRR 5.
FT REPEAT 1202 1225 LRR 6.
FT REPEAT 1227 1248 LRR 7.
FT REPEAT 1249 1271 LRR 8.
FT REPEAT 1273 1295 LRR 9.
FT REPEAT 1314 1337 LRR 10.
FT REPEAT 1338 1360 LRR 11.
FT REPEAT 1361 1384 LRR 12.
FT REPEAT 1386 1407 LRR 13.
FT REPEAT 1408 1430 LRR 14.
FT REPEAT 1432 1455 LRR 15.
FT REPEAT 1459 1529 LRR 16.
FT REPEAT 1533 1556 LRR 17.
FT REPEAT 1557 1580 LRR 18.
FT REPEAT 1581 1604 LRR 19.
FT REPEAT 1606 1628 LRR 20.
FT REPEAT 1633 1656 LRR 21.
FT DOMAIN 1722 2001 PP2C-LIKE.
FT DOMAIN 2002 2493 CATALYTIC.
FT DOMAIN 759 763 POLY-ASP.
FT DOMAIN 888 897 POLY-ALA.
FT METAL 2063 2063 MAGNESIUM (BY SIMILARITY).
FT METAL 2105 2105 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5B58 CRC64;

```
Query Match 3.2%; Score 102; DB 1; Length 2493;
Best Local Similarity 21.7%; Pred. No. 54;
Matches 85; Conservative 44; Mismatches 136; Indels 136; Gaps 17;

QY 4 FTATAVSGRWLGGNHT-----QPLSSQSSDLSYCSLLPMASRVTRKLVSSALHTP 57
DB 801 FKXTASLAGSRGTDSVDLTALPPLGSKSVDEAAANKVDVLOQ-TNNLAQSALVQO- 858

QY 58 PALHFPKQSNPAIVVKPKAKESNTKQNLFORAAALDAEGLVSH-----EKLHPL 113
DB 859 ---QSQSNHQPSPNVRTSRGAGAHM--PASAGAAAAAAGKLGHRPSKRWMMAR 913

QY 114 PXTAD-----PSVQIAGNFAPVNPQVRNLPVVGKLPDSIKGVYVANGANPLHEPTG 167
DB 914 PNTAGSVGATRSSTTLGSLSAED-----TSINGIRDDG-HPLKRSATA 959

QY 168 -----HFFDGDGMHVAKFEHGSASYACRFTQTNRFVQERQLGRPVFPKAIGE 216
DB 960 NTNNATGTLPRNH-----IRVYKTDGTFATLSCLSVSTANEVQ----- 998

QY 217 LHGTCIARLMLFYARAAAGIVDPAGTGVANAGLVYFNGRLAMSEDDLPYQVQ----- 271
DB 999 -----TLARKSTUTTESAAYRLF-----VRDKG-----SERPLGIS--DKPSQLQRRKLI 1041

QY 272 ----ITPNDGLKTVGRDFDQGLESTMIAPHKVPDPSGELFALSVDWSPKYLKYFRFPSPD 328
DB 1042 QAGYTENDGLEDMGRDLS-----YLLRFVFRPD 1070

QY 329 GYKSPDVE-IQDQPTMMDHFAITENFVVVP 358
DB 1071 SVPTFDSIGHSHTFQHLDLHRSNLEMP 1101

RESULT 13
GLGS_BRANA STANDARD; PRT; 520 AA.
AC Q9M462;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate adenyllyltransferase small subunit, chloroplast
DE precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose
DE pyrophosphorylase) (AGPase B) (Alpha-D-glucose-1-phosphate adenyl
DE transferase).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Drakkar; TISSUE=Seed;
RA Zawodny S., Martini N.;
RT "Isolation and analysis of a cDNA clone encoding the small subunit of
RT ADP-glucose pyrophosphorylase in the plastids of seeds and leaves of
RT Oilseed rape (Brassica napus).";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This protein plays a role in synthesis of starch. It
CC catalyzes the synthesis of the activated glycosyl donor, ADP-
CC glucose from GLC-1-P and ATP.
CC -!- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -!- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
CC BY ORTHOPHOSPHATE, ALLOSTERIC REGULATION.
CC -!- PATHWAY: Starch biosynthesis.
CC -!- SUBUNIT: Heterotrimer.
CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -!- TISSUE SPECIFICITY: Leaves.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ271162; CAB89863.1; -
CC InterPro; IPR005836; ADP_Glu_pyrop.
CC InterPro; IPR005835; NTP transferase.
CC Pfam; PF00483; NTP transferase; 1.
CC PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
CC PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
CC PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
CC Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
CC Multigene family; Starch biosynthesis; Allosteric enzyme;
CC Chloroplast; Transit peptide.
CC TRANSIT 1 71 CHLOROPLAST (POTENTIAL).
CC CHAIN 72 520 GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE
CC FT SMALL SUBUNIT.
CC FT SEQUENCE 520 AA; 57044 MW; D0EAF09706F3B6A7 CRC64;
CC SQ
Query Match 3.2%; Score 101.5; DB 1; Length 520;
Best Local Similarity 27.0%; Pred. No. 5.9;
Matches 43; Conservative 27; Mismatches 66; Indels 23; Gaps 7;

QY 1 MASFTATAVSGRWLGGNHTQPLSSQSSDLSYCSLLPMASRVTRKLVSSALHTPPAL 60
DB 1 MATMAAIGSLKVPSSSSNHTRLSSSQSKTLFSFSS---SLTGKLNPTQIILIS--- 53

QY 61 HFPKQSNPAIVVKPKA-KESNTKQNLFORAAALDAEGLV---SHEKLHPL-PK 115
DB 54 NLPRGNERRTPSIVSPKAVSDSONSTCLDPDASRVL---GILLGGAGTRLYPLTKK 109

QY 116 TADPSVQIAGNFAPVNPQVRNLPVVGKLPDSIKGVYV 154
DB 110 RAKPAVPLGANYRLI-----DIPVSNCLNSNISKIYV 141

RESULT 14
SYR_XANCP
ID SYR_XANCP STANDARD; PRT; 562 AA.
AC Q8F455;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-CRNA synthetase (EC 6.1.1.19) (Arginine--trna ligase) (ArgRS).
DE ARGS OR XCC3861.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio A.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
```

```

RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF012507; AAM43092.1; -.
CC HAMAP; MF_00123; -.
CC InterPro; IPR001278; Arg_tRNA-synt_1c.
CC InterPro; IPR005148; N.
CC InterPro; IPR001412; tRNA-synt_1.
CC Pfam; PF03485; N-Arg; 1.
CC Pfam; PF00750; tRNA-synt_1d; 1.
CC PRINTS; PR01038; TRNASYNTHARG.
CC TIGRfams; TIGR00456; args; 1.
CC PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 129 139 "HIGH" REGION.
SQ SEQUENCE 562 AA; 61953 MW; 9750B78D230FB1BD CRC64;

Query Match 3.2%; Score 101.5; DB 1; Length 562;
Best Local Similarity 22.8%; Pred. No. 6.7;
Matches 92; Conservative 48; Mismatches 152; Indels 111; Gaps 20;

QY 71 AIVVKKAKESNTKQMLFORAAALD-----AAEGFLVSH-----EKLHPLPK 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 48 AAMLLAKAASNPALAAQALLALPASDDVARVEIACPGFINFLTPAAYQREVIVIKQ 107
QY 116 TADPSVQIAGNPAFVNEQPVRRN-----LPV-----VGKLPDSIKGVYVRNGANPLHEPTVG 167
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 GHYDYGRLAGNRSVGVYVSANPTGTLPHVGHGRAAAGDSLARVLDAANGWVYGR----- 163
QY 168 HHFEDDGMVHAKFPHGSASYACRTQTNRFOEQLGRFPVPPKAG-ELHGHGTG---- 222
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 --FYNDG--GVQIENLAL-----VQARAQG--LTPDSAGHPENGVRGDYIA 205
QY 223 -IARLML-----FYARAAAGIVDPAGTGVANAGLVYFNGRLAMSEDDLPYQVQITPN 275
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 DVAKAYLAGDVTDLGHLVTGTDPADLESIRRFAYALRNE-----QN 249
QY 276 GDLKTVGRFPDGG-LESTWIAHPKVDPESEGLFAL--SYDVYSKPYLKYFRP----- 325
QY 250 HDLAAP-RVDFDIYFLESSLYKDGKVDKDEAVQKLIASGHTYEEGGALWLKSTDFGDDKDRV 308
QY 326 --SPDGTKS---PDVEQLDQPTMMHDFATE-----NFVYVVPD----- 359
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 MRKSDGYTYFVPDVAVHLTKWQRGYERAITELGADHGSLTRVRAGLQAVELGIPQGW 368
QY 360 QQVVKLPENIRGSGSPVYQKVKVARGILDKVAEDSSN-IKW 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 EYVLHQMVTNMRGGEYKLSKRAGSYVTLRLDLIEETSADAVRW 411
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
SC1 SCHPO STANDARD; PRT; 565 AA.
AC O14335;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-binding protein scr1.
GN SCR1 OR SPBC1D7.02C.
OS Schizosaccharomyces pombe (Fission yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859160;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scourto J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals J.,
RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Farnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN [2]
RP IDENTIFICATION, AND FUNCTION.
RP MEDLINE=98204792; PubMed=9535817;
RA Tanaka N., Ouchi N., Mukai Y., Osaka Y., Ohtani Y., Tabuchi M.,
RA Bhuiyan M.S.A., Fukui H., Harashima S., Takegawa K.,
RT "Isolation and characterization of an invertase and its repressor
RT genes from Schizosaccharomyces pombe."
RL Biochem. Biophys. Res. Commun. 245:246-253(1998).
CC -!- FUNCTION: INVOLVED IN CARBON CATABOLITE REPRESSION. REPRESENTS
CC THE TRANSCRIPTION OF VARIOUS GENES INCLUDING THE INV1 GENE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE CRE/MIG GROUP OF C2H2-TYPE ZINC-
CC FINGERS PROTEINS.
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98270; CAB10978.1; -.
CC PIR; T39863; T39863.
CC HSP; P08047; 1SP2.
CC GenDB_Spombe; SPBC1D7.02C; -.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 2.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00355; Znf_C2H2; 2.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
KW DNA-binding; Transcription regulation; Repressor; Zinc-finger;
KW Nuclear protein; Zinc; Metal-binding; Repeat.
FT ZN FING 26 48 C2H2-TYPE 1.
FT ZN FING 54 78 C2H2-TYPE 2.
FT DOMAIN 81 97 ALA-RICH.
FT DOMAIN 509 548 SER-RICH.

```

```
FT DOMAIN      164      167      POLY-GLN.
FT DOMAIN      375      378      POLY-SER.
SQ SEQUENCE     565 AA; 59713 MW; FE6CCCF7E54A8CDB CRC64;

Query Match
Best Local Similarity 21.8%; Score 101.5; DB 1; Length 565;
Matches 74; Conservative 38; Mismatches 147; Indels 81; Gaps 15;

QY      6  ATAAVSGRWLGGNHTOPPLSSOSSDLS--YCSSLPMASRVTKLNVSSALHTP---PAL 60
Db      147 ATAAVMSY--PHYSASVQQOQATFVSGOPHNLPAAQOPATYIGIPDALLHTQNGTII 204
QY      61  HFPKQSSNGPAIVVKPKAKESNTKQMNLFQRAAAAALDAAGFLVSHEKHLPLPKTADPS 120
Db      205 HVTGTTPPGAVSORSEPDRLSSNMENOLLASAAANQLDAA-----PRIT-PT 250
QY      121 VQIAGNFAPVNEOP-VRRNLPVVGKLPDSIKGYVVRNGANPLH-EPVTGHHFFDGDGMVH 178
Db      251 KSSGVNLMPLUSNAPSPPKQNVVGSIPSS-----SNTSPNHLASVPNRGLTSSNSTGS 303
QY      179 AVKFEHGSASYACRFTQTRRFVQEROLGRPVFP---KAIGELHGHTGIAIARLMLFYARAAA 235
Db      304 FTKYTNGS-----SNSLYSNSMOTPYLPKSKNSSTSLHSMYGVGT----- 344
QY      236 GIYDPAHGIGVANAGLVYFNGRLLAMSEDDLPYQVOITPNGDLKTVGRFDFDQQLSEIMI 295
Db      345 -----TAYAPQSURYAH-----YNYLPYRPSVSGNG-----FDDSSSSDF 380
QY      296 AHPKVDPESGELFALSVDVVSKEPKYKFRFSPDGTKSPDV 335
Db      381 AHFRYQRRSRPVSPCS-TAPSSPTFTSRGSP-----TPDV 415
```

Search completed: November 14, 2003, 21:16:46
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: November 14, 2003, 21:14:17 ; Search time 21 Seconds
(without alignments)
2743.097 Million cell updates/sec
Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGNGHNT.....VPYGFHGTFIGADDLAKQVV 599
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 76.*
1: piri.*
2: piri2.*
3: piri3.*
4: piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2289	72.7	604	T51936	probable 9-cis-epo
2	2280.5	72.4	605	T07123	nine-cis-epoxycaro
3	2170.5	68.9	657	E96812	protein F3P9.10 [i
4	2044.5	64.9	589	A86425	probable 9-cis-epo
5	1991.1	63.2	583	T04531	nine-cis-epoxycaro
6	1930	61.3	604	T04351	viviparus-14 prot
7	968	30.7	595	T04438	hypothetical prote
8	938	29.8	538	T49193	neoxanthin cleavag
9	937	29.7	538	T51734	neoxanthin cleavag
10	745.5	23.7	475	AG1944	hypothetical prote
11	745	23.7	446	T17019	hypothetical prote
12	528.5	16.8	483	E87345	conserved hypothet
13	466	14.8	501	T70534	hypothetical prote
14	454.5	14.4	502	A70582	hypothetical prote
15	394	12.5	490	S76169	hypothetical prote
16	355.5	11.3	497	AE2341	lignostilbene-alph
17	352	11.2	484	JN0595	lignostilbene alph
18	317	10.1	480	S76206	hypothetical prote
19	303.5	9.6	489	JC4324	lignostilbene alph
20	253	8.0	616	T10688	hypothetical prote
21	244	7.7	618	C84885	hypothetical prote
22	242	7.7	472	AG2417	hypothetical prote
23	238.5	7.6	556	F88115	protein F53C3.12 [
24	231	7.3	483	D87290	conserved hypothet
25	215	6.8	533	A47143	retinal pigment m
26	126	4.0	790	T34293	hypothetical prote
27	116.5	3.7	1832	T31113	mucin-like glycopr
28	114	3.6	870	S74291	hypothetical prote
29	113	3.6	466	IQEBV	replication initia

30 110.5 3.5 658 2 S60170 protein kinase Pak
31 108 3.4 437 2 G70019 conserved hypothet
32 107.5 3.4 658 2 T39500 serine/threonine-s
33 107 3.4 1461 2 T13157 mitotic checkpoint
34 107 3.4 26926 1 I38344 titin, cardiac mus
35 106.5 3.4 1299 2 T47182 hypothetical prote
36 106 3.4 1679 2 T15968 chondroitin sulfat
37 105.5 3.3 2591 2 A47171 pristinamycin I sy
38 105 3.3 2504 1 A57788 enoyl-lacyl-carrie
39 105 3.3 2504 1 A57788 MYC-related DNA bi
40 104.5 3.3 623 2 T52293 phycobilisome link
41 104.5 3.3 1080 2 A35088 hypothetical prote
42 103.5 3.3 394 2 AG3616 phenylalanine-trna
43 103 3.3 775 2 B70449 BIR repeat contain
44 103 3.3 4845 2 T31067 cell wall degradat
45 102.5 3.3 531 2 AH3378

ALIGNMENTS

RESULT 1

T51936
Probable 9-cis-epoxycarotenoid dioxygenase [imported] - potato
C:Species: Solanum tuberosum (potato)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T51936
R:Burridge, A.; Taylor, I.B.; Thompson, A.
Submitted to the EMBL Data Library, March 2000
A:Description: Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.
A:Reference number: 225874
A:Accession: T51936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-604 <BUR>
A:Cross-references: EMBL:AJ276244; PIDN:CAB76920.1
C:Genetics:
A:Gene: nced1

Query Match 72.7%; Score 2289; DB 2; Length 604;
Best Local Similarity 71.3%; Pred. No. 1.4e-165;
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;

QY 5 TATAVSGRWLGNGHNTOPPLSSSSQSDLSYCS---SLPMASRVTRKLNVSALHTPPALH 61
DB 3 TTTSHATNTWI-----KPKLSMPSSKEFGFASNSISLLKQHNQSLNINSSLAQPPILH 57
QY 62 FPKOSSN-----SPAIVVKPKAKESNT-----KQMLFORAAAAALDAEGFLVSEKHLPL 113
DB 58 FPKOSSNYQTPTKSTISHPKQENNSSSSISKNLVQKAAAMALDAVEGALTKHELEHPL 117
QY 114 PKTADPSVOIAGNEAPVNEQPVRRNLPVVGKLPDSIKGVVVRNGANPLHPEVPTGHHFFDG 173
DB 118 PKTADPRVQISGNFAPVFNPCQSLPTGKI PKCVQGVVVRNGANPLFEPTAGRHFFDG 177
QY 174 DGMVHAKFEHGSASYACRFQTNRVFOEROLGRVPFPAIGELHGHGTIARLMLFYARA 233
DB 178 DGMVHAVQKNGSASYACRFETERFVQEKALGRVPFPAIGELHGHSGIARLMLFYARG 237
QY 234 AAGIVDPAHGTGVANAGLVYNGRLLAMSEDDL PVQVQITPENGDLKTGVRDFDQQLBEST 293
DB 238 LFLGLDHRGRTGVANAGLVYNNRLLAMSEDDL PVHVKVITPTGDLKTEGRDFDQQLKST 297
QY 294 MIAHPKVPESGELFALSVDYVSKPYLKYPFSPDGTSPDVEIQDQFTMMHDFAITEN 353
DB 298 MIAHPKLPVSGELFALSVDYVQPKYKYPFSPKNGEKSNDEIPEVDPVTMMHDFAITK 357
QY 354 FVVVDDQVVKLPENIRGGSPVVDKKNVAFGLDKYAEKSSNKNIKWIDAPDCFCFHLW 413
DB 358 FVFIIDQVQVFKVSSMIRGGSPVVDKKNVAFGLDKYAKDGGDLKWEVDFDQFCFHLW 417
QY 414 NAWPEPEDEVVIGSCMTPPDSIFNESDENLKVSLSEIRLNKLTGSTRPPIISNEDQQ 473

Db 418 NAWEEPETDEIVVIGSCMTTPDIFNECEDEGLKSVLSEIRLNLKTGKSTRKAIENPDEQ 477
QY 474 VNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKXHLGDNRYGGEPLF 533
Db 478 VNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVEKFIYDKNKYGGEPLF 537
QY 534 LPGE--GGEDEGYILCFVHDEKTKSELQIYNVAVSLEVEATVKLPSRVPYGFHGTFIGA 591
Db 538 LPRDPNSKEBDGYYILAFVHDEKWTSELQIYNVAMTLKLEATVKLPSRVPYGFHGTFIGA 597
QY 592 DDLAKQ 597
Db 598 NDLANQ 603

RESULT 2
T07123
nine-cis-epoxycarotenoid dioxygenase - tomato
N:Alternate names: probable neoxanthin cleavage enzyme
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07123
R:Burbidge, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z15934
A:Accession: T07123
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-605 <BUT>
A:Cross-references: EMBL:297215; PIDN:CAB10168.1

Query Match 72.4%; Score 2280.5; DB 2; Length 605;
Best Local Similarity 71.7%; Pred. No. 6.2e-165;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;
QY 5 TATAAGSRMLGNGHPTPPSSSSQSSDLVCS---SLPMASRVTRKLNYSALHTTPALH 61
Db 3 TTTGHTATWTF-----TKLMSPKSEKGFASNISILLKQNRQSLNINSLSQAPILH 57
QY 62 FPKQSSN--SPA--IVVKPKAKESN-----TKQNLFORAAALDAEGLVSHKELHP 112
Db 58 FPKQSSNVQTPKNNTISHPKQENNNSSSSSTSKNVLQKAAALDAVESALTKHELEHP 117
QY 113 LPKTADPSVQIAGFAPVNEQPVERNLPVVKLPDSIKGVVVRNGANPLPEPVTGHHFFD 172
Db 118 LPKTADPRVQISGNFAPVNPVQCSLPVTGKIPKCVGVVVRNGANPLPEPVTGHHFFD 177
QY 173 GDGMVHAVKFEHGSASACRFTQNRFFVQERQLGRPVFPKPAIGELHGHGTGIARMLFYAR 232
Db 178 GDGMVHAVQPKNGSASACRFTETELVQEKALGRPVFPKPAIGELHGHGTGIARMLFYAR 237
QY 233 AAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLKTVGREDFDGOLES 292
Db 238 GLFGLVDHSHKGTGVANAGLVYFNNRLAMSEDDLPYHVKVPTTGDLKTEGRFDFGQJLS 297
QY 293 TWIAHPKVPDPSGELFALSVDVSKPYLYKFRFPDGTGKSPDVEIQLDQETMMDHFAITE 352
Db 298 TWIAHPKLDPVSGELFALSVDVIQKPYLYKFRFPKNGEKSNDEI PVEDPTMMDHFAITE 357
QY 353 NFVVVPDQVQVFKLPEMIRGSGSPVYDKNVARGILDKYAEEDSSNKNWIDAPDPCFPHL 412
Db 358 NFVWIPDQVQVFKMSEIRGSGSPVYDKNVSRFGILDKYAKGSDLKWVEVPDPCFPHL 417
QY 413 NNAWEEPEDEVVVIGSCMTTPDPSIFNESDENLKSLSVSEIRLNLKTGSTRRPPIISNEQ 472
Db 418 NNAWEEATEDEIVVIGSCMTTPDPSIFNECEDEGLKSVLSEIRLNLKTGKSTRKAIENPDE 477
QY 473 QVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKXHLGDNRYGGEPL 532
Db 478 QVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVEKFIYDKNKYGGEPL 537
QY 533 FLPGS--GGEDEGYILCFVHDEKTKSELQIYNVAVSLEVEATVKLPSRVPYGFHGTFIG 590

Db 538 FLPRDPNSKEBDGYYILAFVHDEKWKSELQIYNVAMSLKLEATVKLPSRVPYGFHGTFIG 597
QY 591 ADDLAKQ 597
Db 598 ANDLANQ 604

RESULT 3
E36812
protein F399.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E36812
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E36812
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <STO>
A:Cross-references: GB:AB005173; NID:g8052533; PIDN:AAF71797.1; GSPDB:GN00141
C:Genetics:
A:Gene: F3F9.10
A:Map position: 1

Query Match 68.9%; Score 2170.5; DB 2; Length 657;
Best Local Similarity 67.7%; Pred. No. 1.6e-156;
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;
QY 1 MASFATATAVAGRMV-----GNHTOPPLSSSSQSSDLVCSLSP-----MASRVT 45
Db 51 MASFTLLSTSTQFLDRFTSSSSSRPKQS-----LSFSTLRNKKLVPCVSSSVN 105
QY 46 RKLNVSSALHT-----PPALHFFKQSSNSPAIVVVKPKAKESNTKQNLFORAAALDAE 101
Db 106 KKSSVSSLSQSTFKPPSW---KKLCNDVTNLI-PKTTNQNP-KLNPVQRTAAVMDAVE 160
QY 102 GLVSHK-KLHPLPKTADPSVQIAGNAPVNEQPVRRNLPVVKLPDSIKGVVVRNGANP 160
Db 161 NAWISHERRRRPHKPTADPAVQIAGNFPFVEKPVVHNLPTGTVPECIQGVVVRNGANP 220
QY 161 LHEPVTGHHFFDGDGMVHAVKFEHGSASACRFTQNRFFVQERQLGRPVFPKPAIGELHGH 220
Db 221 LHKPVSGHHLFDGDMGVHAVRFDNGSVSYACRFTETNRLVQERECGRFVFPKPAIGELHGH 280
QY 221 TGIARMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLKT 280
Db 281 LGIAKLMLFNTKGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLPVHVKTQTGDLLET 340
QY 281 VGRFDPDGOLESTMIAPKVPDPSGELFALSVDVSKPYLYKFRFPDGTGKSPDVEIQLD 340
Db 341 SGRYDFDGLQKXSTMIAPKIDPETRELFALSVDVSKPYLYKFRFTSDGEKSPDVEIPLD 400
QY 341 QPTMMDHFAITNFVVPDQVQVFKLPEMIRGSGSPVYDKNVARGILDKYAEEDSSNKN 400
Db 401 QPTMIDHFAITNFVVPDQVQVFKLPEMIRGSGSPVYDKNVARGILDKYAEEDSSNKN 460
QY 401 WTDAPDCFCFHLNNAWEEPEDEVVVIGSCMTTPDPSIFNESDENLKSLSVSEIRLNLKTGE 460
Db 461 WLEVPDPCFCFHLNNSWEEPEDEVVVIGSCMTTPDPSIFNEHDETTLSVLSSEIRLNLKTGE 520
QY 461 STRPPIISNEQOVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKH 520
Db 521 STRPPIISNEQOVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKH 577

521	LYGDNRYCGBELFLFPGEGEEDBQYILCFVHDEKTKWSELQINVANSLEVEATVKLPSRV	580
QY		
578	LYGEGYKGGEFLFPGSGDEGGYIMVFHDBEKVKQLINAVNMKDEATVTLPSRV	637
Db		
581	PYGFHGTFIGADDLAKQVV	599
QY		
638	PYGFHGTISKEDLSKOAL	656
Db		

RESULT 4

A86425
 probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana
 C-Species: Arabidopsis thaliana (mouse-ear cress)
 C-Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C-Accession: A86425
 R-Authors: Thaler, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A-Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzilli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A-Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A-Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A-Reference number: A86141; MUID:12016719; PMID:11130712
 A-Accession: A86425
 A-Status: preliminary
 A-Molecule type: DNA
 A-Residues: 1-589 <STO>
 A-Cross-references: GB:AE005172; NID:g11094779; PIDN:AAG29711.1; GSFDB:GN000141
 C-Genetics:
 A-Map position: 1

Query Match	64.9%;	Score 2044.5;	DB 2;	Length 589;
Best Local Similarity	68.1%;	Pred. No. 4.9e-147;		
Matches 390;	Conservative 69;	Mismatches 101;	Indels 13;	Gaps 4
QY	27	SOSSDLSTYCSSLPMASSRVTRKLNVSALHTPPALPHFKQSSNSPAIVWPKAKESNTQM	86	
DB	27	SPSSVSFTNTKPR-----RRKLSANSVSDTPTNLNLFVNYPSNPPII-----PEKOTS	76	
QY	87	NLFQRAAAAALDAAGFLVLSHEKHLPIKPTADPSQVAGNFAPVNPQPVARNLPVVGKLP	146	
DB	77	NPLQRAASAALDFAETALLRRERSKLPKPTVDPHQVSGNYAPVPEQSVKSSLSVDGKIP	136	
QY	147	DSIKGVYVANGANPIHPEVPTGHHFFDGDGMVHAVKVFHGHGSASVACRPTQTNRRPVQBRQIG	206	
DB	137	DCIDGVLNLANGANPLFEPVSGHILFDGDMVHAVKITNGDASYSCRPTETERLVQSKQLG	196	
QY	207	RVPFPKATGELCHGTGIARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLLAMGEDDI	266	
DB	197	SPIFPKAIGELGHGSGIARLMLFYARGLFGLNHNKGTGVANAGLVYFHDRLAMGEDDL	256	
QY	267	PYQVQITPFGDLTKTVGRDFDQGLSTIAHPKVDPSGELFALSVDWSKPYLKYFRFS	326	
DB	257	PQVVRVTDNGDLLETTGRDFDQGLSSAMIAHPKIDPVTKELFALSVDVVKPYLYFKFS	316	
QY	327	PQGTSPDVEIOLDQFTWMHDPATENTENVVVDQOVFKJPEMIRGGSPVYDKNKVAAP	386	
DB	317	PEGEKSPDVEIPLASFTWMHDPATENTENVVTPDQOVFKLSDFMLGKSPVKYDGEKISR	376	
QY	387	GILDKYAEPSDSNIKVIDAPDCCFHLNNAWBPETDEWVVGSCWTPPDSIFNESDENLK	446	
DB	377	GILPRNAKADASMWVSPETFCFHLNNAWESPETDEVVVGSCWTPPADSIFNECDEQLN	436	
QY	447	SVLSEIRLNLTKTGSTRPITISNEQQVNLNLAGMVRNMLGRKTKFAYALALAPWPVKVG	506	
DB	437	SVLSEIRLNLTKGSTRITIPG-SVQVNLNLAGMVRNMLGRKTRAYALALAPWPVKVG	495	
QY	507	FAKVDLTTCEVVKHLYGDNRRYGEZPLFPG--EGGESEDEGVILCFVHDEKTKSEIQVNL	564	
DB	496	FAKVDLSTCEVKNHFFYGGKYGCEPFFLPRLGESGDDGYFMFVHDEGSWESELIHVN	555	

Qy	565	AVSLEVEATVKLSRPVPGFHGTFIGADDLAKQ	597
	: :		
Db	556	AVTLEAEATVKLSRPVPGFHGTFSADMLNQ	588

RESULT 5

T04531
n9n-cis-epoxycarotenoid dioxygenase homolog F28J12.10 - Arabidopsis thaliana
N:Alternate names: hypothetical protein T9A21.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04531; T04937
R:Bavan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, K.F.
submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15377
A:Accession: T04531
A:Molecule type: DNA
A:Residues: 1-583 <BEV>
A:Cross-references: EMBL:AL021710
A:Experimental source: cultivar Columbia; BAC clone F28J12
R:Bavan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15390
A:Accession: T04937
A:Molecule type: DNA
A:Residues: 1-377 <BSW>
A:Cross-references: EMBL:AL021713
A:Experimental source: cultivar Columbia; BAC clone T9A21
C:Genetics:
A:Map position: 4
A:Note: F28J12.10; T9A21.200

Query Match	63.2%	Score	1991	DB 2	Length	583
Best Local Similarity	64.5%	Pred. No.	5.5e-143	Indels	22	Gaps
Matches	380	Conservative	82	Mismatches	105	
Qy*	16	GGNHTQPLSSQSSDLSY--CSSLPWASRVTRKLNLYSSALHTPPALHPFKQSSNSPAIV	73			
D5	11	GGIKTWP-----QAQIDLGFRPIKQPKVICKTQVIDVTE-LTKKRLQFLTPRTATTP---	62			
Qy	74	VKPKAKESNYQMNLFORAAAAALDAAGELVSHKEKHLPLKPTADPVSQIAGNAPVNEQ	133			
D5	63	-----QHNPLRLNI FQAAAIAIDAERALISHEQDSPLKPTADPRVQIAGNYSVPVES	116			
Qy	134	PVRNLPVVGKLPDSISIKGVYVRNGANPLHEPVTGHEFFDGGMVHAKVFEHGSASYACRF	193			
D5	117	SVRENLTVEGTPICIDGVYIRNGANPMWFEPTAGHLLFDGGMVHAKVITNGSASYACRF	176			
Qy	194	TQTRFVQEROLGRPVFPKALGELHGTGIARLMLFYARAAGIVDPAGHTGVANAGLVY	253			
D5	177	TKTERLVQERKLGPRVPFPAIGEHLHSGGIARLMLFYARGLCGLINNGVGVANAGLVY	236			
Qy	254	FNGRLAMSEDDLFPYQVQITPENGDLKTVGRFPDGOLESTMAHPKVPDPESGELFALSVD	313			
D5	237	FNNRLAMSEDDLFPYQLKITQTDGLQTVGRVDFDGLKSAIMAHPKLPDVTKELHALSYD	296			
Qy	314	VVSKPYLYKYPFSSDGTKS PDVEITLOOPTMMHDFAITENFVVPDQOVFKLPEMIRGG	373			
D5	297	VVKXPYLYKYPFSSDGVKSPSELEIPLETPTMIHDFAITENFVLPDQOVFKLGEIMISGK	356			
Qy	374	SPVYVDKNKVARFGLDKYASDSSNIKWIDAPDCFPHLMAWEEPETDEVVYVIGSCMTP	433			
D5	357	SPVYVDGEKVSRGLGMPKDATEAQSIWNSPETFCPHLMAWESPETEIVVIGSCMSP	416			
Qy	434	PDLSIFNESDENLKSVLSEIRLNLTGSTRPILISNEDQQVNLEAGVNRNMLGRKTRFA	493			
D5	417	ADSLFNEKDESLRSLSEIRNLNTRKTKTRSLLVNED--VNLEIGMVRNRLGRKTRFA	474			
Qy	494	YLALAEPMPKVSGFAKVDLTITGVEKKHLYGDNRYGGEFLFUPGEGG----BEDEGYILCF	549			
D5	475	FLATAYPMPKVSGFAKVDLTGCEMKKYYIGCEKYGGEFFYPGNSCNGEENEDDGYISCH	534			

[illegible]

T04438

hypothetical protein T18B16.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04438
R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Bancroft, I.; Mewes, H.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15359
A:Accession: T04438
A:Molecule type: DNA
A:Residues: 1-595 <REV>
A:Cross-references: EMBL:AL021687
A:Experimental source: cultivar Columbia; BAC clone T18B16
C:Genetics:
A:Map position: 4
A>Note: T18B16.140

Query Match 30.7%; Score 968; DB 2; Length 595;
Best Local Similarity 36.7%; Pred. No. 2,5e-65;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVSSALHTPPALHFPKQSNSGPAIVVKPKAKESNTKQNLFORAAAAALDAE 101
DB :
DQ SPITNPSDDNRNKPKTLF---NRNTHTLVSSPPKLPEMTLATALF---TTVEDVIN 94
QY 102 GLVLVSHEKLHLPLKPATDSQVAGNFAPVNEOPVRRLNPV-GKLPDSIKGYVVANGANP 160
DB :
DQ 95 TFIDP-----PRSPSVDPKHVSDNFAPVLDELPTDCDEIHGTLSINGAYIRNGPNP 149
QY 161 LHPVPTHGFDFDGGMVAHKVEHGASVACHFTQTNRVQRQLGRPVFFKAIGELHGH 220
DB :
DQ 150 QFLPRGPYHLFDGDGMLHAIKHNGKATLCRSRYVTYKNYEKOTGAPVMNVFGFNQV 209
QY 221 T-GIAELMLFYARAAAGIYDPAHGTGVANAGLVYFNGRLAMSEDDLPHYVOITPNGDLK 279
DB :
DQ 210 TASVARGALTAAVLTGQYNPVNGICLANTSLAFFSNRLFALGESDLLPYAVRLTESGDIE 269
QY 280 TVGRFDFDGOLESTWTAHKVPDPESGELPALSYDVVSKPYLYKYFRSPDGTKSPDVEI-Q 338
DB :
DQ 270 TIQRYDFDGLKMSMTAHPKTPITGETFAFYGFV-PPELFYFPDSAGKKQROVPIFS 328
QY 339 LDQPTWMHPFAITENFWVPDQQWFK---LPEMIRGGSPVYVDXKNVARFGILDKYAE 395
DB :
DQ 329 MTSPSFTHDFAITKRHAIFAEQLGWRNNMLDLVLEGSGVGTDNKTPTRLGVIPKYAGD 388
QY 396 SSNIKVIDADPCFCFLHNAWEPEPTDEVVVGSCWTPDPSIFNESDENLKSLVELRLN 455
DB :
DQ 389 ESEMKKFEVPGFNIIHAINAWEDDGNISVLIAPNIMSIEHTLERMD-LVHALVEKKVID 447
QY 456 LKTGESTRRRIISNEDQQVNLEAGMVNRNMLRKTKTFAYLALAEWPVKVSGFAKVLDLTG 515
DB :
DQ 448 LVTVGVRHHPI SAR-----NLDPAVINPAFLGRCSRYVYAAIGDPWKISGVVVKLDVSKG 502
QY 516 E-----VKKHLYGDNRYGSEPLPDEGG-----ERDEGYLLCFVHDEKTKWSQLQVNAV 566
DB :
DQ 503 DRDCTVAREMYSGCYGGEFFVARDPGNPEAEDDGYVTVYVHDDEVTSGESKFLVMDAK 562
QY 567 S--LEVEATVKLPSRPYPYGFHGTFFIGADDLAK 596
DB :
DQ 563 SPELEIVAVALPRRPYPYGFHGI-FVKESDLNK 594

RESULT 8
T49193
neoxanthin cleavage enzyme ncl - Arabidopsis thaliana
N:Alternate names: protein MA21.150
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49193
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25018
A:Accession: T49193

RESULT 7

RESULT 8

T49193
 Nixanthin cleavage enzyme ncl - *Arabidopsis thaliana*
 Alternate names: protein MAA21.150
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #t
 C:Accession: T49193
 R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25018
 A:Accession: T49193

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-538 <RIE>
A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150
A:Experimental source: cultivar Columbia; BAC clone MAA21
C:Genetics:
A:Gene: ATSP:MAA21.150
A:Map position: 3 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49
A:Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4.1e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIVVVKPK-AKESNTKQMLFQRAAAAALDAEGFLVSHEKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIISVHPRPSKGFSSKLLDLERLVVKLM-----HDSLEPLH-----Y 47

QY 123 IAGNFAPV-NEQPVRRNLFWGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVK 181
DB 4 KLSGSSIIISVHPRPSKGFSSKLLDLERLVVKLM-----HDSLEPLH-----Y 47

QY 48 LSGNFAPIRDETTPVKDLFVHGFLPECLNGEFVRVGNPKFQDAVAGYHWFDDGMHIGVR 107

QY 182 FHGGSASYACRTQTNRFVQERQOLGRPVFPKPAIGELHGTGIARLMLFYARAAAGIVDPA 241
DB 108 IKDGKATYVSRYVYKTSRLKQEBEFGGAAPKMK-IGDLKGFGLLVNVOQLRTKILNT 166

QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVOITPNDGKLVGRFDFDQGLSESTMIAPKVD 301
DB 167 YNGTANTALVYHGHKLLAQEADKPVYIKVLEDGLQTLGIIDYDKRLTHSFTAHPKVD 226

QY 302 PESGELFALSYDVVSKPYLYKFRFSPDGTSKSDVEIQLODPTMMHDFAITENFVVVDDQ 361
DB 227 PVTGEMFTFGYS-HTPPYLTIRVISKDGIMHDPVITISEPIMHDFAITETVYAFMDLP 285

QY 362 VVFKLPEMIRGGSPV-YDNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPE 420
DB 286 MHPREKMWKEMKMIYSFDPKARFGLPRYAKDELMIKWFELNCFIHNANAWEE-E 343

QY 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEQQVN 475
DB 344 EDEVLITCRLENPDLDMVSGVKLEFNENFNMKTSASQKLSASA----- 398

QY 476 LEAGMVRNMLGRKTFAYLALAEPPKVSQFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTGKQKRVYVGTILDSIAKVTGIIFDLHAEAEKGMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFPGEGEEDGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPRETAEDDGYLIFFVHDENTGKSCVTVIDAKTWSAEPVAVVELPHRV 518

QY 581 PYGPHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLEQTL 537

RESULT 9
T51734
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51734
R:Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.
J. Exp. Bot. 49, 1893-1894, 1998
A:Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture
A:Reference number: Z24454
A:Accession: T51734
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-538 <NEI>
A:Cross-references: EMBL:AJ005813; PIDN:CAA06712.1
A:Experimental source: cultivar Landsberg erecta
C:Genetics:
A:Gene: ncl

Query Match 29.7%; Score 937; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4.8e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIVVVKPK-AKESNTKQMLFQRAAAAALDAEGFLVSHEKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIISVHPRPSKGFSSKLLDLERLVVKLM-----HDSLEPLH-----Y 47

QY 123 IAGNFAPV-NEQPVRRNLFWGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVK 181
DB 48 LSGNFAPIRDETTPVKDLFVHGFLPECLNGEFVRVGNPKFQDAVAGYHWFDDGMHIGVR 107

QY 182 FHGGSASYACRTQTNRFVQERQOLGRPVFPKPAIGELHGTGIARLMLFYARAAAGIVDPA 241
DB 108 IKDGKATYVSRYVYKTSRLKQEBEFGGAAPKMK-IGDLKGFGLLVNVOQLRTKILNT 166

QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVOITPNDGKLVGRFDFDQGLSESTMIAPKVD 301
DB 167 YNGTANTALVYHGHKLLAQEADKPVYIKVLEDGLQTLGIIDYDKRLTHSFTAHPKVD 226

QY 302 PESGELFALSYDVVSKPYLYKFRFSPDGTSKSDVEIQLODPTMMHDFAITENFVVVDDQ 361
DB 227 PVTGEMFTFGYS-HTPPYLTIRVISKDGIMHDPVITISEPIMHDFAITETVYAFMDLP 285

QY 362 VVFKLPEMIRGGSPV-YDNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPE 420
DB 286 MHPREKMWKEMKMIYSFDPKARFGLPRYAKDELMIKWFELNCFIHNANAWEE-E 343

QY 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEQQVN 475
DB 344 EDEVLITCRLENPDLDMVSGVKLEFNENFNMKTSASQKLSASA----- 398

QY 476 LEAGMVRNMLGRKTFAYLALAEPPKVSQFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTGKQKRVYVGTILDSIAKVTGIIFDLHAEAEKGMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFPGEGEEDGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPRETAEDDGYLIFFVHDENTGKSCVTVIDAKTWSAEPVAVVELPHRV 518

QY 581 PYGPHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLEQTL 537

RESULT 10
AG1944
hypothetical protein all1106 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1944
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Kaneko, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, D.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073063.1; PID:gl7130452; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1106

Query Match 23.7%; Score 745.5; DB 2; Length 475;
Best Local Similarity 34.5%; Pred. No. 1.4e-46;
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

QY 123 IAGNFAPVNEQPVRRNLFWGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVK 182
DB 48 LSGNFAPIRDETTPVKDLFVHGFLPECLNGEFVRVGNPKFQDAVAGYHWFDDGMHIGVR 107

Db 22 LEQNFAPVBBEITDTTLKIGELPPELSGMFVRNGPNPQWPTPIGOYHWFDDGQMLHGVRI 81
QY 183 EHGSAVACRPTQNRVQEROLGRVPFPAIGALHGHGTGIAKLMLFYARAAAGIYD--- 239
Db 82 SNGKATYNRVVRQRIEHAQAIAW-----TGLVEPQ 117
QY 240 ---PAHGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDDGQLESTMA 296
Db 118 TELSPRNTG--NTALIWHAGOLLALWEGGAPYATQVP---DLASIGEYTYNNQLSSAFTA 172
QY 297 HPKVDPESEGFALSVDVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVV 356
Db 173 HPKVDPTVGEWMPFGYS--FAPPYLHYSVSATGELVTVTIDLPWGMWMDFAITANYTI 231
QY 357 VPDQOVFKLEPMIRGGSPVYDKNKVARFOILDKYABDSNNKIMWIDAPCFCHLWNAW 416
Db 232 FMDLPLTFVSVMQGBMLMFESDRSPRFGILPRHG--DNSQIRWFEPAPSCYVFHTINAY 230
QY 417 EEPETDEVVIG-----SCTMPDPSIFNESDENLK--SVLSEIRLNKLTGSTRRPIISN 469
Db 291 E--DKDEVVLFPACPMRSTTVLASPDS---QTDPEADIPRLHRWFHKLTKGVHEEVL--- 342
QY 470 EDQOVNLEAGMVRNMLGRKTKFAYLA--LAE--PWPKVSGFAKVDLTGTVKVKHLYGDNRY 527
Db 343 --DVASEFPRIENFLQPTQYGVTSRLAKGSIPLFEGLIKYLDSNAKSQNYEYQGRY 400
QY 528 GGEPLFLPGEGB--EDEGYILCFVHDEKTKWSELOIVNAVSLVSE--ATVKLPSRVYPYGF 584
Db 401 GSEAVFVRPRGATVEDDGLNLTIVYDTEBSESELVIVINAQDINSEPIARVLLPORVYPGF 460
QY 585 HGTTIGADDL 594
Db 461 HGIWVTEBQL 470

RESULT 11
T17019
Hypothetical protein - apple tree
C:Species: Malus domestica (apple tree)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T17019
R:Watillon, B.; Kettmann, R.; Aredouani, A.; Hecquet, J.; Boxus, P.; Burny, A.
Plant Mol. Biol. 36, 909-915, 1998
A:Title: Apple messenger RNAs related to bacterial lignostilbene dioxygenase and plant
A:Reference number: Z18655; MUID:98179104; PMID:9520281
A:Accession: T17019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-446 <WAT>
A:Cross-references: EMBL:293765; NID:G2924324; PIDN:CAR07784.1; PID:G2924325
A:Experimental source: cv. McIntosh, strain Wjczk

Query Match 23.7%; Score 745; DB 2; Length 446;
Best Local Similarity 39.8%; Pred. No. 1.3e-48;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;
QY 231 ARAAGIVDPAGHTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDDGQ 290
Db 76 ASVLTVGQYNPANGIGLANTSLAFAFGDRLYALGESDLPYSLRSLTNGDIETLGRHDFDGL 135
QY 291 ESTMTAHKVPDSESEGFALSVDVSKPYLKYFRFSPDGTKSPDVEI--QLDQPTMMHDF 349
Db 136 SNMNTAHPKIDPTGEAFRAGFI--RPLTYFRFDSNGVKQPDVPISWVTPPTFLHDF 194
QY 350 ITENFVVVDDQOVFKLEPMI--RGGSPVYDKNKVARFGILDKYABDSNNKIMWIDAPCF 408
Db 195 ITKGAIFAQIQLNLDIMTKRATPGLDPSKVPRIQVIPLYAKDESKRWFEVPGFN 254
QY 409 CFHLNNAWEPETDEVVWVIGSOWTPDPSIFNESDENLKSLSSEIRLNKLTGSTRRPIIS 468
Db 255 GVHATNWD--EDDAIVWVAPNVLSEHVLERVD--LVHCLVEKVAIDLKTVGIVTEQLST 311
QY 469 NEDQOVNLEAGMVRNMLGRKTKFAYLAEPWPKVSGFAKVDLTGTG-----VKHLYG 523

Db 312 R-----NLDFAVINPAYLGRNKYVYAAGDPMKISGVVVKLDVSNVEHKECIVASRMFG 366
QY 524 DNRYGGEPLFLPGE---GGEDEGYILCFVHDEKTKWSELOIVNAVSLVSE--ATVKLPSRVYPYGF 577
Db 367 PCYGGGEFFVAREPENPEADEDNGFLVSYVHDEKAGESRFLVMDAKSPQLDIIAAVRMP 426
QY 578 SRVYGFHGTGIGADDL 594
Db 427 RRVYGFHGLFVRESDL 443

RESULT 12
E87345
conserved hypothetical protein CC0776 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87345
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AE005673; NID:gl3422015; PIDN:AAK22761.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0776

Query Match 16.8%; Score 528.5; DB 2; Length 483;
Best Local Similarity 30.9%; Pred. No. 4.1e-32;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;
QY 108 EKLHLPKTDPS--VOIAGNFAPVNEQPVNRNLPVW--GKLPDSIKGVYVNGANPLHEP 164
Db 4 ERLPPVRLSLGTNHPYMTGPMTPQHEEVNANWLELEGAIPADLDGVLRTENPVDHP 63
QY 165 VTGHHFFDGDGMVHAKVFEHGSAYACRFTOTNRFVQERQLGRPVFPKAI GELHGHGTGIA 224
Db 64 IGRYHFFDGDGMHIEFKGGAATYRNFVTRCFEAEQVNEGLW---GGLMDGPGVS 119
QY 225 RLMLFVARAAAGIVDPAGHT--GVANAGLVFNGRLLAMSEDDLPYQVQITPNDGL 278
Db 120 K-----RFGFG---AHGALKDSASTDIVVHNGEAIATP-----YQGEAYRUDPL--TL 163
QY 279 KTVGRFPDQGLESTMIAPKVDPSGELFALSVDVWSK--PYLKYFRFSPDGTKSPDVE 336
Db 164 ENLGVASW--APLEG--VSAHPKYDEATGELMFPNY---SKAWPYMHYGVVGPDKRKYQG 218
QY 337 IQLDQPTMMHDFAITENFVVVDDQOVFKLEPMIRGGSPVYDKNKVARFGILDKYABDS 396
Db 219 VPLPGRLPHDVAFSKYAILNCLFVFWDQELMARDIHAVRLHKGIPSRFALVPR---EG 275
QY 397 SNIKWIDAPDCFCFLHNAWEPETDEVVWVIGSOWTPP-----DSIFNESDEN 444
Db 276 GEPRWEAEPTVYLWLNAYE--DGDEVLDGYFQEKIPRELEGAPDGHGHLMAYLDEH 333
QY 445 LKSVLSEI---RLNKLTGSTRRPIISNEDQOVNLEAGMVRNMLGRKTKFAYLAEP- 500
Db 334 --SFLPKLHRWRPNLTKGTETTEKHL-----DDRVL--LEFGMFKQYAGKYRAYSSTAKPG 386
QY 501 WPKVSGFAKVDLTGTVKVKHLYGDNRYGGERFLPQEGG--EDEGYILCFVHDEKTKWSE 559
Db 387 WFLFNGFVKHDLTETGSNSIALPEGYASEAPFAPKVGAVDEDDGVLVSFIIDENKAGE 446
QY 560 LOIVNAVSLVEATVKLPSRVYPYGFHGTGIGADDLAK 596
Db 447 CLIVDAKREVEVCRIALPHKLSSGCTHVWAGREMLTK 483

Search completed: November 14, 2003, 21:18:12
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2003, 22:18:27 ; Search time 516 Seconds
(without alignments)

3793.904 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAAAGRWLGNGHT.....VPYGHGTFIGADLLAKQVV 599

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2169961 segs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ .p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09758269/runat_14112003.192311.25851/app_query.fasta.1.775
-DB=Published Applications NA -QWMT=fastap -SUFFI=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORW=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 USER=US09758269@cgn 1 1 347 @runat 14112003 192311.25851
-NCFU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3150	100.0	1800	10	US-09-758-269-5	Sequence 5, Appli

2	2280.5	72.4	1818	10	US-09-758-269-15	Sequence 15, Appl
3	2168.5	68.8	1839	10	US-09-758-269-11	Sequence 11, Appl
4	1991	63.2	1752	10	US-09-758-269-1	Sequence 1, Appl
5	1930	61.3	1815	10	US-09-758-269-13	Sequence 13, Appl
6	1663.5	52.8	1734	10	US-09-758-269-9	Sequence 9, Appl
7	969	30.8	1788	10	US-09-758-269-3	Sequence 3, Appl
8	969	30.8	1788	10	US-09-938-842A-1444	Sequence 1444, Ap
9	937	29.8	1617	10	US-09-758-269-7	Sequence 7, Appl
10	937	29.7	1617	10	US-09-758-269-17	Sequence 17, Appl
11	522	16.6	393	10	US-09-878-574-2543	Sequence 2543, Ap
12	419	13.3	32679	8	US-08-976-0630-21	Sequence 21, Appl
13	411.5	13.1	1518	8	US-08-976-0630-21	Sequence 21, Appl
14	369	11.7	320	10	US-09-878-574-2872	Sequence 2872, Ap
15	346	11.0	230	9	US-09-770-696-172	Sequence 172, App
16	279	8.9	534	12	US-10-149-753-65	Sequence 65, Appl
17	260	8.3	3111	14	US-10-053-192-2	Sequence 2, Appl
18	238.5	7.6	2134	12	US-10-168-517-18	Sequence 18, Appl
19	234.5	7.4	367	9	US-09-770-791-402	Sequence 402, Appl
20	232.5	7.4	1855	12	US-10-168-517-16	Sequence 16, Appl
21	207.5	6.6	1934	12	US-10-168-517-20	Sequence 20, Appl
22	150	4.8	2037	12	US-10-168-517-1	Sequence 1, Appl
23	133	4.2	82993	15	US-10-080-170-645	Sequence 645, App
24	127.5	4.0	13029	9	US-09-815-242-4052	Sequence 4052, Ap
25	122	3.9	5127	14	US-10-156-761-3630	Sequence 3630, Ap
26	122	3.9	9025608	14	US-10-156-761-1	Sequence 1, Appl
27	111.5	3.5	1875	14	US-10-156-761-3143	Sequence 3143, Ap
28	111	3.5	1956	14	US-10-156-761-4583	Sequence 4583, Ap
29	110	3.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
30	108.5	3.4	4146	12	US-10-032-585-6784	Sequence 6784, Ap
31	107	3.4	2825	12	US-10-289-757-15	Sequence 15, Appl
32	107	3.4	3075	9	US-09-833-790-424	Sequence 202, App
33	107	3.4	7515	12	US-10-101-510-202	Sequence 4, Appl
34	107	3.4	7515	12	US-10-210-120-4	Sequence 1, Appl
35	107	3.4	81940	10	US-09-759-508B-1	Sequence 724, App
36	107	3.4	81940	12	US-09-873-319-724	Sequence 1092, Ap
37	107	3.4	81940	12	US-09-360-706-1092	Sequence 7396, Ap
38	106.5	3.4	1083	9	US-09-815-242-7396	Sequence 82, Appl
39	106.5	3.4	10174	14	US-10-171-311-82	Sequence 76, Appl
40	106	3.4	1239	10	US-09-922-261-76	Sequence 75, Appl
41	106	3.4	1825	10	US-09-922-261-75	Sequence 2792, Ap
42	106	3.4	2514	14	US-10-156-761-2792	Sequence 1, Appl
43	106	3.4	4308	12	US-10-102-549-1	Sequence 67, Appl
44	106	3.4	4331	12	US-10-354-358-67	Sequence 3146, Ap
45	105.5	3.3	3048	14	US-10-156-761-3146	

ALIGNMENTS

RESULT 1
US-09-758-269-5
; Sequence 5, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (1)...(1797)
US-09-758-269-5

Alignment Scores:

Pred. No.: 0 Length: 1800
Score: 3150.00 Matches: 599
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-758-269-6 (1-599) x US-09-758-269-5 (1-1800)

```
QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
Db 1 ATGGCTTCTTTCACGGCAACGGCTGCGGTTCTTGGAGATGGCTTGGTGGCAATCACT 60
QY 21 GlnProPheLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 61 CAGCGCGCATATCGCTTCTTCMAAGCTCCGACTTGAGTTATTGTAGTCCCTTACCTATG 120
QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeu 60
Db 121 GCCAGTCGTGCACACGTAAAGTCAATGTTTCATCTCGCTTCACACTCCCTCCAGCTCTT 180
QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
Db 181 CATTTCCCTTAAGCAATCATCAAACTCTCCGCCCATTTGTTAAGCCCAAGCCAAAGAA 240
QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
Db 241 TCCAACTAAACAGATGAATTTGTTCCAGAGCGCGCGCGAGCGGTTGGACGGCGCG 300
QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
Db 301 GAGGGTTTCCTTGTTCAGCCACGAGAGCTACCCCGCTTCTTAAACGGGTGATCTTAGT 360
QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
Db 361 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCGCTCCGCGGTAACTTCGG 420
QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db 421 GTGGTCGGAACACTTCCCGATTCATCAAGAGGTGATGCGCAACGGAGCTTAACCCA 480
QY 161 LeuHisGluProValThrGlyHisPheAspGlyAspGlyMetValHisAlaVal 180
Db 481 CTTTCAGACCGCGTGACAGGTCAACACTTCTTCGACGAGACGGTATGTTTCACGCGTC 540
QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db 541 AAATTTCGAACACGGTTCAGTACGCTTACCGCTTTCGCCGTTTACTCAGACTAACCGGTTGT 600
QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db 601 CAGGAACTGTCANTTGGGTGACCGGTTTTCGCCAAGCCATCGGTGACCTTCACGGCCAC 660
QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
Db 661 ACCGGTATTGCCCGACTCATGCTATTCTACGCGCAGAGTGCAGCGGTATAGTCGACCGG 720
QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
Db 721 GCACCGGAACCGGTGACTAACCGCGGTTTGGTCTATTTCATAGCGCGGTTATAGTGGCT 780
QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
Db 781 ATGTCGGAGATGATTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAAAACC 840
QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
Db 841 GTTGGTGGGTTCGATTTCATGACAAATTAGAAATCCCAATGATTTGCCACCCGGAAGTC 900
```

```
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
Db 901 GACCCGGAATCCGGTGAACCTTTCGCTTAAAGTACAGCTGCTTCAAGACCTTACCTA 960
QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
Db 961 AAATACCTCCGATTCTCACCGGACGGAATAAATCAACCGGACGTCGAGATTCAGCTTAT 1020
QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360
Db 1021 CAGCAACAGATGATGCACGATTTCGCGATTACAGAGAACTTCGTCGTCGACCTGACAG 1080
QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
Db 1081 CAATCGCTTTCAGCTGCCGAGATGATCCGGGTGGTCTCCGGTGGTTTACGACAAG 1140
QY 381 AsnLysValAlaAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db 1141 AACAGGTCCGCAAGATTTCGGGATTTCAGCAAAATACCGCAAGATTTCATCGAATTAAG 1200
QY 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
Db 1201 TGGATTGATGCTCAGATTGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGACCCAGAA 1260
QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
Db 1261 ACAGATGAAGTCGTCGATAGGCTCTGTATGATCCACCAGACTCAATTTTCAACGAG 1320
QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db 1321 TCTGACGAGATCTCAAGAGTGTCTGCTGAAATCCGCTGAATCTCAAAACCGGTGAA 1380
QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
Db 1381 TCAACTCCGCGTCCGATCATCTCCAACGAAGATCAACAAAGTCAACCTCGAAGCAGGATG 1440
QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
Db 1441 GTCAACAGAAACATGCTCGCCGTAAACCAATTCGCTTACTTGGCTTTAGCCGAGCGG 1500
QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrGlyGluValLysLysHis 520
Db 1501 TGGCTTAAAGTCTCAGGATTCGTAAGTTCATCTCACTTCTGGAGAAAGTTAAGAAACAT 1560
QY 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGlyGlyGlyGlu 540
Db 1561 CTTTACGGCGATAACCGTTTACGAGAGAGAGCTCTGTTCTCCCGGAGAGGAGAGAG 1620
QY 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db 1621 GAAGACGAAGGATACATCTCTGTTTCGTTCCAGCAGAGAACATGGAATTCGAGTTA 1680
QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db 1681 CAGATAGTTAAGCCCTTAGCTTAGAGTTGAAGCAACGGTTAACTTCCGTCAGGGTT 1740
QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db 1741 CCGTACGGATTCACCGTACATTCATCGGAGCCGATGATTTCGCGAAGCAGGTCGTG 1797
```

RESULT 2

US-09-758-269-15
; Sequence 15, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI SATOSHI
; APPLICANT: KOBAYASHI, KAZUO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12

```

; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1815)
; US-09-758-269-15

Alignment Scores:
Pred. No.: 4,34e-252 Length: 1818
Score: 2280.50 Matches: 435
Percent Similarity: 82.04% Conservative: 63
Best Local Similarity: 71.66% Mismatches: 90
Query Match: 72.40% Indels: 19
DB: 10 Gaps: 6

US-09-758-269-6 (1-599) x US-09-758-269-15 (1-1818)
QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyValSerHisThrGlnProLeu 24
D 7 ACTACTACTTCATGCCCAATATACATGATTT-----AAGACTAAGTTG 51
QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41
D 52 TCAATGCCATCATCAAGAGGATTTGGTTTGGCATCAAACTATTCTCTACTCAAAAT 111
QY 42 SerArgValThrArgLysLeuValSerSerAlaLeuHisThrProAlaLeuHis 61
D 112 CAACATAAATAGGCAAAAGTCTCAACATAATTCCTCTCTCAAGCTCCACTATATCTAT 171
QY 62 PheProLysGlnSerSerAsn-----SerProAla-----IleValValLysProLys 77
D 172 TTTCTCTAAACAATCTTCAAAATATCAACACCAACCAAGAAATAATACAAATTTTCAACACCAAAA 231
QY 78 AlaLysGlnSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92
D 232 CAAGAAACCAACACACTCTCTCTTCAACTTCCAAAGTGGAAATTTAGTCGAAAGCA 291
QY 93 AlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112
D 292 GCAGCAATGCTTTAGATGCTGTGAGAAGTCTTTAACTAAACATCAACTTGAACACCCCT 351
QY 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132
D 352 TTGCCGAAACACGCCGACCCACGAGTCCAGATTTCTGGGAATTTTGTCCGGTACCGGAA 411
QY 133 GlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal 152
D 412 AATCCAGTCTGTCATCTCTCCGGTCCACCGGAAATAATACCCAAATGTGTTCAAGCGTT 471
QY 153 TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePheAsp 172
D 472 TACGTTTCCAAACCGGAGCTAACCTCTTTTTCACCAACACCCCGGACACCATTTCTTCGAC 531
QY 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg 192
D 532 GGCACCGGTATGTTTACGCGGTCAATTCAAAATGGGTGGGTAGTACGTTGCCGT 591
QY 193 PheThrGlnThrAsnA-gPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212
D 592 TTCCTGAAACAGAGAGGCTTGTTCAGAAAAAGCTTTGGGTGCCCTGTTTCCCTAAA 651
QY 213 AlaIleGlyGlnLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArg 232
D 652 GCCATTGTGTAATTACATGGTCACTCTGGAATTCGAAGGCTTATGCTGTTTACGCTCGT 711

```

```

QY 233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuVal 252
D 712 GGGCTCTTCGGACTTGTGATCAGATAAAGAACTGGTGTTCGAAACCGCGGTTAGTC 771
QY 253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272
D 772 TATTTCAATAACCGATTACTTGTCTGTAAGATGATTTGGCTTACCATGTAAAGSTA 831
QY 273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292
D 832 ACACCCACCGCGCATCTTAAACAGAGGGTTCGATTTCGATTCACCGCCAGCTAAATCC 891
QY 293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312
D 892 ACCATGATAGTCTACCCAAAGTCGACCCAGTTTCCGGTGAGCTATTTGCTCTTAGCTAC 951
QY 313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332
D 952 GATGTGATTCAGAGCGCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCA 1011
QY 333 ProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGlu 352
D 1012 AATGATGTTGAATTCAGTTTGAAGACCCCAACATGATGATGATTCGCAATTACTGAG 1071
QY 353 AsnPheValValProAspGlnValValPheLysLeuProGluMetIleArgGly 372
D 1072 AACTTCGTCGTCATCTGATCAACAGTCTGTTTCAAGATGCTGAAATGATCCCGTGA 1131
QY 373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392
D 1132 GGTTACCGGTGGTTTACGACACAGACAAAGTTTCCGATTTGGTATTCTGGATAGTAC 1191
QY 393 AlaGluAspSerSerAsnIleLysTyrIleAspAlaProAspCysPheCysPheHisLeu 412
D 1192 GCGAAAGATGGTCTGATTTGAAATGGTTCAGAGTCACTGATTTGTTCTGTTCCACCTC 1251
QY 413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432
D 1252 TGGATGCTTGGGAAGAGCAGAAACAGATGAAATCGTTTGAATTTGTTTATGTTATGATGACA 1311
QY 433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIle 452
D 1312 CCACCACTCCATTTTCAATGATGATGATGAGGGCTTAAAGAGTGTGTTTATCCGAATC 1371
QY 453 ArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleSerAsnGluAspGln 472
D 1372 CGTCTCAATTTGAAAAACAGGAAATCAACAGAAATCCATAATCGAAAAACCCCGATGAA 1431
QY 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492
D 1432 CAAGTGAATTTAGAGCTGGAATGTTGACCGAAACAACTCGGAGGAAAAACAGAGTAT 1491
QY 493 AlaTyrLeuAlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAspLeu 512
D 1492 GCTTATTTGGGTATCGGTGAAACCATGGCCAAAAGTTTCTGGTTTTCGAAAGTAAACCTG 1551
QY 513 ThrThrGlyGluValLysLysHisLysTyrGlyAspAsnArgTyrGlyGluProLeu 532
D 1552 TTCACCGGTGAAGTTGAGAAATTCATTTATGTTGTCACCAACAATATGTTGGGGAACCTCTT 1611
QY 533 PheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIleLeuCysPheVal 550
D 1612 TTTTACCAAGAGACCCCAACAGCAAGAGAGACGATGTTTATATTTTACGTTTTCGTT 1671
QY 551 HisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570
D 1672 CACGATGAGAAGAAATGGAATTCAGACTGCAATTTGTTAACGCAATGAGTTTGAAGTTG 1731
QY 571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590
D 1732 GAGGCAACTGTGAAGCTTCCATCAAGAGTTCCTTATGATTTCTATGATTTTCATGGAACATTCAATAC 1791
QY 591 AlaAspLeuAlaLysGln 597

```

Db 1792 GCCAATGATTGGCAATCAG 1812
||||:|||||
RESULT 3
US-09-758-269-11
; Sequence 11, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: KOBAYASHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1839
; TYPE: DNA
; ORGANISM: Vigna unguiculata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1836)
US-09-758-269-11
Alignment Scores:
Pred. No.: 3,53e-239 Length: 1839
Score: 2168.50 Matches: 423
Percent Similarity: 78.44% Conservative: 61
Best Local Similarity: 68.56% Mismatches: 98
Query Match: 68.84% Indels: 35
DB: 10 Gaps: 8
US-09-758-269-6 (1-595) x US-09-758-269-11 (1-1839)
QY 8 AlaAlaValSerGlyArgTyrLeuGlyGlyAsnHisThrGlnProLeuSerSerSer 27
:||||:|||||
Db 7 TCATCAGCTTCAACACTGGTTAAAGCCACACTCCCATCTCCCTCCCTTCAAGACCTA 66
:||||:|||||
QY 28 GlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLys 47
:||||:|||||
Db 67 CCTTCCACATCT--TCTCCCAACAACTTACTTCTTTA-----AGAAAAACATCTCT 117
:||||:|||||
QY 48 LeuAsn---ValSerSerAlaLeuHisThrProAlaLeuHisPheProLysGln--- 65
:||||:|||||
Db 118 TCCACACACCATCACAATGCTCCCTTCAACA-----CTCCACTTCCCAACACATAC 169
:||||:|||||
QY 66 -----SerSerAsnSerProAlaLeuValValLysProLysAlaLysGluSer 81
:||||:|||||
Db 169 CAACCAACATCCATCCATCCACAGCCACCACCAACACACCCCAATCAAAACT 228
:||||:|||||
QY 82 AsnThr----- 83
:||||:|||||
Db 229 ACCACCATCACCACCAACACCGCCAGGAAACCAACCTCTCTGTACACCAACCAA 288
:||||:|||||
QY 84 -----LysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
:||||:|||||
Db 289 CCATTACCTCAAAATGGAACTTCTCCAGAAAGCGCTGCCCGCTTGGACCTGGTC 348
:||||:|||||
QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
:||||:|||||
Db 349 GAAACGGCGCTCGTCTCGCAGCGCAACACCGCTCCCAACACCGCGGACCCGAGG 408
:||||:|||||
QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
:||||:|||||
Db 409 GTCCAAATCGCCGGGAACCTCGCGCGGTGCGGAGCATGCCCGCATCAAGGACTCCCG 468
:||||:|||||
QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
:||||:|||||
Db 469 GTGGTCGGAATAATCCCAATGCAATGACGGCGGTGTACGTGCGCAACGGTGCATCCG 528
:||||:|||||
QY 161 LeuHisGluProValThrGlyHisHisPheAspGlyAspGlyMetValHisAlaVal 180
:||||:|||||
Db 529 CTCTACGAGCTGTGGCGCGGACCACTTCTCGAGCGGACGCGCATGGTCCACGCGGTG 588
:||||:|||||
QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
:||||:|||||
Db 589 AAGTTCAGAACGCGCGCGCGCTACGCTGCGCGTTCACCGAGAGCGACGCGTCTCTCG 648
:||||:|||||
QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
:||||:|||||
Db 649 CAGGAGAAATCTCTAGCGCGCGCGGTGTCTCCGAGAGCGCATCGGGAGAGTCCACGCGCAC 708
:||||:|||||
QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
:||||:|||||
Db 709 TCCGCGATCGCGCGCTCTCTCTTACGCGCGCGGTCTTTCGGGCTCGTTGATGGG 768
:||||:|||||
QY 241 AlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAla 260
:||||:|||||
Db 769 TCCGAGGCGATGGCGGTGGCGAAGCGCGGTCTGCTTCTTCAACACCAACCTCTTGCC 828
:||||:|||||
QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
:||||:|||||
Db 829 ATGTCGGAAGACGATTTTACCTTACCACTGAGAAATCACCCCTAACCGCGCATTAACACC 888
:||||:|||||
QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
:||||:|||||
Db 889 GTTGCGCGTTACGACTTCAAGCGGCGCTCAACTCAACATGATCGCCCAACCTCTGCGC 948
:||||:|||||
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
:||||:|||||
Db 949 GACCCCGTCGACGCGGACCTCCAGCGCTCAGCTACGAGCTCATTCAGAACCTTACCTC 1008
:||||:|||||
QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
:||||:|||||
Db 1009 AAGTACTTCCGTTTCTCCCGCGCGCTCAAGTCCCGCGCTGGAATCCCTCTGAAG 1068
:||||:|||||
QY 341 GlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
:||||:|||||
Db 1069 GAGCCCAACCATGATCAGCATTTCCGCATAACGAGAAATTTCTGCTGCTCTCCCGACAG 1128
:||||:|||||
QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
:||||:|||||
Db 1129 CAGGTGGTCTTCAAACTAACGAGATGATCACCGCGGGTCCCCCGTGTCTACGACAAG 1188
:||||:|||||
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnLys 400
:||||:|||||
Db 1189 AACAAACCTCAGGTTTGGGATTCTGCACAAAGATGCGAAGACGCGAATGCGATGCGG 1248
:||||:|||||
QY 401 TrpIleAspAlaProAspCysPheCysPheHisIleuTyrAsnAlaTrpGluGluProGlu 420
:||||:|||||
Db 1249 TGGATCGACGCGCGGATTGTTTCTGCTTCCACCTCTGAAACCGCTGGAGAGGCCGAA 1308
:||||:|||||
QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
:||||:|||||
Db 1309 ACCGAGGAGTTGTGGTATGGTCTGTCATGACCCCTGCGCATCTCCATTTTCAACGAA 1368
:||||:|||||
QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLysThrGlyGlu 460
:||||:|||||
Db 1369 TGGCAGGAGAGTTGAAGACGCTGTCTCAGAGAAGGCTGAATTTGGGACCGGCAAG 1428
:||||:|||||
QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
:||||:|||||
Db 1429 TCCACTCGCGCGCCCATTTATCTCC---GACCGCGAACAACTGAACCTGGAAGCGGCATG 1485
:||||:|||||
QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
:||||:|||||
Db 1486 GTCAACAGAAACAAAGCTCGGAAGGAAACCCAGTTTCGCTATCTGCTCTGCGGAGCCCC 1545
:||||:|||||
QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
:||||:|||||

Db 1546 TGGCCCAAGTCTCGGGCTTCGGAAGTGTGTTGCTGAGTGGGAAGTGAAGAGTAC 1605
Qy 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlu 540
Db 1606 ATGTATGGAGAGAGAGATTCGGTGGGAGCCTCTGTTCTTCC--AAAGGCCAAAA 1662
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTyrLysSerGluLeu 560
Db 1663 GAAGACGATGGGTATATCTGCATCTGTCACGACGAGAAAGATGGAAATCCGAGCTG 1722
Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db 1723 CAGATGTGAATGCCCAAAATTAAGCTCGAAGCTTCATCAAACTCCCTCTCGTGT 1782
Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597
Db 1783 CCTACGGTTCATGGAACCTTCATTCATTCGAAGATTTCAGGAACAA 1833

RESULT 4
US-09-758-269-1
; Sequence 1, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758, 269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1749)
US-09-758-269-1

Alignment Scores:
Pred. No.: 9.18e-219 Length: 1752
Score: 1991.00 Matches: 380
Percent Similarity: 78.44% Conservative: 82
Best Local Similarity: 64.52% Mismatches: 105
Query Match: 63.21% Indels: 22
DB: 10 Gaps: 6

US-09-758-269-6 (1-599) x US-09-758-269-1 (1-1752)
Qy 16 GlyGlyAsnHisThrGlnProLeuSerSerSerGlnSerSerAspLeuSerTyr--- 34
Db 31 GGTGGTATTAAACATGGGCT-----CAAGCCCAATGATTTGGGTTTAGG 78
Qy 35 ---CysSerSerLeuProMetAlaSerArgValThrGlyLeuAsnValSerAla 53
Db 79 CCCATTAAGAACACCAAGAGGTATTAAATGCACGGTGCAGATCGACGTAACGGAA--- 135
Qy 54 LeuHisThrProAlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleVal 73
Db 136 TTAACCAAAAACGCCAATATTATACCCAGAACCCCGCTACTCCCGCG----- 186
Qy 74 ValIleProLysAlaLysGlnMetAsnThrLysGlnMetAsnLeuPheGlnArgAla 93
Db 187 -----CAGCATATCTCTCCGGCTAAACATCTTCCAGAAACGCGG*228

Qy 94 AlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeu 113
Db 229 GCATTTGCATTCACGCGCTGAGCGTGCATTAACTCAGAGAGCAAGATTTCCACTT 288
Qy 114 ProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGln 133
Db 289 CCCAAACCGCTCATCGTGTTCAGATTGCGGGAATTATTCCCGGTCACCGAATCT 348
Qy 134 ProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyr 153
Db 349 TCCGTCGCGGAAACCTCACCGTCGAAGGAACATCCCTGACGTGATTACGGTGTTAT 408
Qy 154 ValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePheAspGly 173
Db 409 ATCCGTAACCGCGGAATCCGATTTGAGCCAAACAGCTGGGCACCATTTATTCGACGA 468
Qy 174 AspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPhe 193
Db 469 GACGGAATGGTTCACGAGTTAAATAACACCGGTTACGTAGCTACCGATGCCGTTT 528
Qy 194 ThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAla 213
Db 529 AAAAAACCGAGAGATTGGTTTCAGGAAAAACGATTTGGTTCGACCATTTTCCGAAAGCA 588
Qy 214 IleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAla 233
Db 589 ATCGGCGAGCTTCACGTCACCTCGGAATCGCACGTTTGTGCTGTGTGTACGACGTGG 648
Qy 234 AlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyr 253
Db 649 CTTTGTGCTGATCAACACCAACACCGCGCTCGGAGTAGCAACCGCTTGTTGTTTAC 708
Qy 254 PheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThr 273
Db 709 TTTAATAACCGGCTTTAGCTATGTGAGAGACGATTTACCGTACCATAATAAAATTA 768
Qy 274 ProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThr 293
Db 769 CAACCGCGCATCTCCAAACCGTTGACGTTAGCATTCGACGTCAGTTAAATCGCA 828
Qy 294 MetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAsp 313
Db 829 ATGATAGCTCACCCGAACCTGGACCGGTTACGAGGAGGCTTCACGCGTTAAGCTACGAC 888
Qy 314 ValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerPro 333
Db 889 GTGTTAAGAAACCTTACCTGAAATTAATTCAGATTTCGCCAGACGCGGTTAAATCGCCG 948
Qy 334 AspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIleThrGluAsn 353
Db 949 GAATGGAGATCCCGCTCGAAACTCGACGATGATTTCAGATTTCCGTATACCGAGAAAT 1008
Qy 354 PheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGlyGly 373
Db 1009 TTTGTGTGATTCTGTATCAACAGTCGTGTTCAAGCTCGCGAGATGATTTCCGGTAA 1068
Qy 374 SerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAla 393
Db 1069 TCTCCGTTGTTTCGACGCGAGAAAGGTTTCCCGATTGGGGATAATGCCAAGGACGCG 1128
Qy 394 GluAspSerSerAsnIleLysTyrIleAspAlaProAspCysPheCysPheHisLeuTyr 413
Db 1129 ACAGAACCTTCTCAGATAATCTGGGTGAATCTCCGAGACGCTGTCTTTCTATCTCTGG 1188
Qy 414 AsnAlaTyrGluProGluThrAspGluValValIleGlySerCysMetThrPro 433
Db 1189 AATGCATGGGAATCCCGGAGACGAGAGATTGTGGTATCGATCGATCGTGTATGTCGCG 1248
Qy 434 ProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArg 453
Db 1249 GCGGATTCATCTTCAACGAGAGACGAGAGCTTGAGAGCGCTTTGTGCGAGATCAGG 1308
Qy 454 LeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGln 473

```

Db      1309 ATAAACCTCAGAACCGTAAACCGCGCTGCTCGTTGTTGTTAAACGAGAT----- 1362
Qy      474 ValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAla 493
Db      1363 GTAAATTTAGAGATTGTTATGTTAAACCGACCGGTTAGGAAGAAACCCGGTTTCGG 1422
Qy      494 TyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThr 513
Db      1423 TTTTGGCTATTGCTTATGCTTGGCCAAAAGTTTCGGTTTCGCTTAAGTTCGATCTTTCG 1482
Qy      514 ThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPhe 533
Db      1483 ACCGGTGAATGAAATAATATTTACGGCGGTGAGAAATATGGCGGAACCGTTTTC 1542
Qy      534 LeuProGlyGluGlyGly-----GluGluAspGluGlyTyrIleLeuCysPhe 549
Db      1543 TTCCCGCGCACTCCCGTAACCGCGAAGAAATGAAGATGACGGTTATATATTTTGTAC 1602
Qy      550 ValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu 569
Db      1603 GTTCATGACGAAGAAACAAGACATCAGAGCTTCAGATTATTAAACGCTGTTAATTAAAG 1662
Qy      570 ValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIle 589
Db      1663 CTTGAAGCTACGATTAACTACCGTCTAGAGTACCGTATGGGTTTCATGGCACATTGTG 1722
Qy      590 GlyAlaAspAspLeuAlaLysGlnVal 598
Db      1723 GATTGAATGAACCTCGTTGATCAATTA 1749

```

RESULT 5

```

US-09-758-269-13
; Sequence 13, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1812)
US-09-758-269-13

```

Alignment Scores:

```

Pred. No.: 1,03e-211 Length: 1815
Score: 1930.00 Matches: 375
Percent Similarity: 75.68% Conservative: 73
Best Local Similarity: 63.34% Mismatches: 122
Query Match: 61.27% Indels: 22
DB: 10 Gaps: 8

```

US-09-758-269-6 (1-599) x US-09-758-269-13 (1-1815)

```

Qy      22 ProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMetAla 41
Db      49 CCGCGCGGTCACGCGCGCGGCGCTCC-----AATCCGTCAGGTTCTCG 93

```

```

Qy      42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrPro-----Pro 58
Db      94 CCGCGCGCGCTCAGCTCCGTCGCGCGCGCGAGTGCCTCAGCGCGCTTCCACAGCC 153
Qy      59 AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleValVal----- 74
Db      154 GTCCGCGACCTGCTCGCGCTCCAGAAAGCCCGCCCATTCGCTCCAGGCGACGCC 213
Qy      75 ---LysProLysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAla 93
Db      214 CCGCGCGCGAGAAAGCGGAGCGCGCAAGACAGCTCAACTTGTTCAGCGCGCGCG 273
Qy      94 AlaAlaAlaLeuAspAla---AlaGluGlyPheLeuValSer---HisGluLysLeuHis 111
Db      274 CCGCGCGCGCTCGACGGTTCGAGGAAGGGTTCGTGCCCAACGCTCCTCGAGCGGCC 333
Qy      112 ProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsn 131
Db      334 GGGTGTCCACAGCAGCGCGCGCGCGCGTGCAGATCGCGCAACTTCGCGCGCGCGG 393
Qy      132 GluGlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGly 151
Db      394 GAGAGCGCGCGGTGCACGAGCTCCCGCTCCGCGCGCATCCGCGCTTCATCGAGGG 453
Qy      152 ValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePhe 171
Db      454 GTCTACCGCGCAACCGCGCGCAACCCCTGCTTCGACCCCGTCCGCGCGCGCACCTTTC 513
Qy      172 AspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAla---SerTyrAla 190
Db      514 GACGCGACGGCATGTTGCGCGTGCAGCGTGCAGTACGGAACGCGCGCGCGAGTCTTACGCC 573
Qy      191 CysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhe 210
Db      574 TCGCGCTTACGGAGACCGCGCGCTCGCGCAGAGCGCGCGATCGCGCGCGCTTTC 633
Qy      211 ProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyr 230
Db      634 CCCAAGGCCATTGGCGAGCTGCACGGGACTCGCGGATCGCGCGCTCGCCCTTGTCTAC 693
Qy      231 AlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGly 250
Db      694 CCGCGCGCGCGTGCAGCGCTCGTGACCCCTCGCGCGGACCGCGCGTGGCGCAACCGCGG 753
Qy      251 LeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnVal 270
Db      754 CTCGTCTACTTCAACGCGCGCTGCTCGCCATGTCGAGGACGACCTCCCTTACCAGCTC 813
Qy      271 GlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeu 290
Db      814 CCGCTGGCGGACGACGCGGACCTCGACCGCTCGCGCGCTACGACTTCGACGGGACGCTC 873
Qy      291 GluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeu 310
Db      874 GGCTGGCGCCATGATCGCGCACCCCAAGTCGAGACCCCGCGGAGCTCCACGCGCTC 933
Qy      311 SerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThr 330
Db      934 AGCTACGACGTCATCAAGAGCGCGTACTCAAGTACTTCTTACTTCAGCGCGCGACGCC 993
Qy      331 LysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIle 350
Db      994 AAGTCCGACGACGCTGGAGATCCCGCTGGAGCAGCCACGATGATCCACGACTTCGCCATC 1053
Qy      351 ThrGluAsnPheValValProAspGlnValValPheLysLeuProGluMetIle 370
Db      1054 ACCGAGAACTTCTGGTGTGTGCCCGACCCAGCGTGTGTTCAGAGCTCCAGGAGATGCTG 1113
Qy      371 ArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAsp 390
Db      1114 CCGCGCGGCTCGCGCGTGTGTCGCAAGGAGAGACGTCGCGGTTTCGCGGCTCTCCCC 1173
Qy      391 LysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPhe 410

```

```
Db 1174 AAGCAGCGCGGACGCGTCGAGATGCGTGGGTGGAGCGCGGAGCTGTTCTGCTTC 1233
Qy 411 HisLeuTrpAsnAlaTrpGluProGluThrAspGluValValIleGlySerCys 430
Db 1234 CACCTGTGGACGCGTGGAGAGCGGACGCGGAGGTGGTGTGATCGGCTCTGC 1293
Qy 431 MetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSer 450
Db 1294 ATGACCCCGCGGACTCCATCTCAACAGTCCACGAGCGCTGGAGCGGTGCTGACC 1353
Qy 451 GluIleArgLeuAsnLeuLysThrGluSerThrArgArgProIleIleSerAsnGlu 470
Db 1354 GAGATCCCGCTGGAGCGCGGACGCGGCGGTCCAGCGCGCGCGCTCCGCGCGG 1410
Qy 471 AspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThr 490
Db 1411 TCGCAGCAGGAGAACCTGAGGTGGCATGTGTGAACCGCAACCTGCTGGCGCGGAG 1470
Qy 491 LysPheAlaTrpLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysVal 510
Db 1471 CGGTACGGGTACCTCGCGGTGGCGAGCGCGGTGGCCCAAGGAGTGGGCTTCGCG 1530
Qy 511 AspLeuThrThrGlyGluValLysLysHisLeuTrpGlyAspAsnArgTrpGlyGlu 530
Db 1531 GACCTGTCCACGGGCGACTCACCAAGTTCAGTACGGCGAGGCGCGGTTCGCGCGG 1590
Qy 531 ProLeuPheLeuProGlyGluGlyGlyGlu-----GluAspGluGlyTyr 545
Db 1591 CCCTGCTCGTTCCTCCATGACCGCGCGCGCGCCACCGCGCGCGGAGGACGCGGTAC 1650
Qy 546 IleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAla 565
Db 1651 GTGCTCACTTCGTCCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1710
Qy 566 ValSerLeuGluValGluAlaThrValLysLeuProSerArgValProTrpGlyPheHis 585
Db 1711 GCGCAGATCCGCGTGGAGCGCACGCTTCAGTGGCGTCCCGCGTGGCGCTTCGCT 1770
Qy 586 GlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597
Db 1771 GGCACCTTCATCAGCGGCGAGGAGCTCGAGCCCGAG 1806
```

RESULT 6

```
US-09-758-269-9
; Sequence 9, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: KOHCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1731)
US-09-758-269-9
```

Alignment Scores:

Pred. No.: 4,836-181 Length: 1734

```
Score: 1663.50 Matches: 323
Percent Similarity: 72.56% Conservative: 100
Best Local Similarity: 55.40% Mismatches: 137
Query Match: 52.81% Indels: 23
DB: 10 Gaps: 9

US-09-758-269-6 (1-599) x US-09-758-269-9 (1-1734)

Qy 24 LeuSerSerSerGlnSerSerAspLeuSerTrpCysSerSerLeuProMetAlaSerArg 43
Db 25 CTTCTCCGACGACGATCTCTCTGTTCTCTATTCTCCACAAACCCAAAAATGCAAAAT 84
Qy 44 ValThrArgLysLeuAsnValSerSer---AlaLeuHisThrProProAlaLeuHisPhe 62
Db 85 ATTCTCGACGAAATCTCATTAACCTTTCAGAGTACGACACATCTCTGATCTCTACTTCT 144
Qy 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGluSerAsn 82
Db 145 CCGGTTCCGTCACCG-----GTTAAGCTCAAACCAACGATCCAAAC 186
Qy 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaGlu--- 101
Db 187 -----TTAAACCTTCTCAGAGCTAGCGGTACGATGCTCGCAAGATTGAGTCC 237
Qy 102 GlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerVal 121
Db 238 TCTATCGTTATTCCTATCGAGCAGAAATCGCCGCTTCTTAACCCGACCGCGCGT 297
Qy 122 GlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProVal 141
Db 298 CAATTACAGTAACCTTCGCTCCGTTAATGAATGTCGGTTTCAGAACGGTTTGAAGTG 357
Qy 142 ValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyValAsnProLeu 161
Db 358 GTTGTGCAATCTCTTCTTAAAGAGTTTACATCCGTAAACGGTGCAACCCCTATG 417
Qy 162 HisGluProValTrpGlyHisPheAspGlyAspGlyMetValHisAlaValLys 181
Db 418 TTTCCGCGCTTAGCGGACATCATTTATTGACGGTACGGAATGATTCACGCCGTAGT 477
Qy 182 -----PheGluHisGlySerAlaSerTrpAlaCysArgPheThrGlnThrAsnArgPhe 199
Db 478 ATCGGTTTTGATAAC---CAGGTTAGTTACAGCTCCGCTACACTAAACAAACCGGCT 534
Qy 200 ValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyLeuHisGly 219
Db 535 GTTCAAGAAACCGCGCTTCGACGATCGGTTTTTCCCTAAACCAATCGGCGAGCTTCAC 594
Qy 220 HisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAsp 239
Db 595 CATTCGCGTCTAGCTGACATCGCTCTCTTCACGGCTCGAGTGGGATCGGTCTAGTGGAC 654
Qy 240 ProAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeu 259
Db 655 GGGACACGTGGCATGGCGGTAGCTAACGCGGTGGTGGTTTCTTTAAACGGCAGGTTATTA 714
Qy 260 AlaMetSerGluAspAspLeuProTrpGlnValGlnIleThrProAsnGlyAspLeuLys 279
Db 715 GCCATGTCAGAGATGATCTTCTTACCAGTGAAGATCGACGGTCAAGGAGATCTTGAG 774
Qy 280 ThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLys 299
Db 775 ACGATCGGACGGTTCGATTTCGATGACCGAGTTGACTCTTCAGTGTATGCGCATCTTAAG 834
Qy 300 ValAspProGluSerGlyGluLeuPheAlaLeuSerTrpAspValValSerLysProTyr 319
Db 835 GTGGACCGCACACAGGAGATCTCCATACACTGAGCTACACGTTTGAAGAAACCTCAT 894
Qy 320 LeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeu 339
Db 895 CTCAGGTATCTTAATTCACACGTCGCGGAAAAAGACACGTCAGCTGGAGATCAGCTC 954
Qy 340 AspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAsp 359
```


629 CGCGCTCAGTACGTCGTGAGCTTTAAACGGCAGCTAGGTTTAAACCGCAGCATATAATC 688
QY 240 rAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyPheAsnGlyArgLeuLeuA 260
Db 689 CGGTTAAACGGCATGGTTAGCTAATACAAAGTCTAGCTTCTTCAGTAACCGTCTCTTTG 748
QY 260 laMerSerGluAspLeuProTyGlnValGlnIleThrProAsnGlyAspLeuLysT 280
Db 749 CTTTAGGTGAATCTGATTACCTACCGCTCCGATTAACCGAATCAGGAGATATTGAA 808
QY 280 hrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysV 300
Db 809 CGATCGGACGGTACGATTTCGACGGGAATACGATGAGTATGACAGCTCACTCTAA 868
QY 300 alAspProGluSerGlyGluLeuPheAlaLeuSerTyAspValValSerTyProTyL 320
Db 869 CCGATCCAATAACCGGAGAACTTCGCTTCCTCGGTACGGTCCGGT---CCACCGTTT 925
QY 320 euLysTyPheArgPheSerProAspGlyThyLysSerProAspValGluIle---GlnL 339
Db 926 TAAACATATTTCCGGTTTGGATTCGCCCGGGAACAAAGACAGCTTCGGATATTCTCGA 985
QY 339 euAspGlnProThrMetHisAspPheAlaIleThrGluAsnPheValValProA 359
Db 986 TGAGCTCTCGTCTCTCCATGACTTCGGATCACGAAACGTACGCGATTTCGCGAC 1045
QY 359 spGlnGlnValPheLys-----LeuProGluMetIleArgGlyGlySerProV 376
Db 1046 AGATTACGCTTGGCATGAGATGAACATGTTGGATTGGTTCTCGAAGGTGGTTCTCCGG 1105
QY 376 alValTyAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyAlaGluAspS 396
Db 1106 TTGGTACTGATACCGGAAAACTCCAGGCTTGGAGTATCTCAAGTACCGCGGAGATG 1165
QY 396 erSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaT 416
Db 1166 AGTCGGAGATGAATGGTTCCGAGTCTCGATTCTCAATATCATTCACGCTATTAAATGCTT 1225
QY 416 rpGluGluProGluThrAspGluValValIleGlySerCysMetThrProAspS 436
Db 1226 GGGATGAAGATGATGGAACAGCGCTGTTTGTATGTCACCGAATATTATTCGATTCGAC 1285
QY 436 erIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnL 456
Db 1286 ATACTTTAGAGAGATGGAT---CTGGTTTCATGCTTTGGTGGAGAGTGAAGATCGATC 1342
QY 456 euLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnValAsnL 476
Db 1343 TCGTCACCGGATTTGTGAGAGTCAATCCGATCTCAGCGAGG-----AATC 1387
QY 476 euGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyLeuA 496
Db 1388 TCGATTTCGCTGTGATTAAATCCGGCTTTCTCGGAGATGTAGCAGGTACGTTTACGCGG 1447
QY 496 laLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyG 516
Db 1448 CGATTGGAGATCCGATCCGGAAGATCTCCGTTGGTGCAACTTGATGTGTCTAAAGGAG 1507
QY 516 lu-----ValLysLysHisLeuTyGlyAspAsnArgTyGlyGlyGluP 531
Db 1508 ATCCGGATGATTGTACGGTGGCCCGCTAGAAATGATACGGTTACGGTTGTTACGCGGAGA 1567
QY 531 rLeuPheLeuProGlyGluGlyGly-----GluGluAspGluGlyTyIleL 547
Db 1568 CGTTTTTCGACTAGGATCTCGTAAATCCGGAGCGGAGGAGGATGATGTTATGTGG 1627
QY 547 euCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValS 567
Db 1628 TGACGTATGTTTACGATGAAGTACTGGAGAAATCGAAGTTTCTGTGTATGCGCGTAAT 1687
QY 567 er-----LeuGluValGluAlaThrValLysLeuProSerArgValProTyGlyPheH 585
Db 1688 CGCCGGAGGTTGAAATCGTCGCCCGGTGAGTTCGCCGGAAGGGTTCGATACGGATTC 1747

QY 585 isGlyThrPheIleGlyAlaAspAspLeuAlaLys 596
Db 1748 ATGGGTTAATTGTCAAGGAAAGTGACCTTAATAAG 1782
RESULT 9
US-09-758-269-7
; Sequence 7, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1614)
US-09-758-269-7
Alignment Scores:
Pred. No.: 1,64e-97 Length: 1617
Score: 938.00 Matches: 207
Percent Similarity: 55.81% Conservative: 105
Best Local Similarity: 37.03% Mismatches: 199
Query Match: 29.78% Indels: 48
DB: 13 Gaps: 13
US-09-758-269-6 (1-599) x US-09-758-269-7 (1-1617)
QY 64 LysGlnSerSerAsnSerProAlaIleValValLysProLys---AlaLysGluSerAsn 82
Db 10 AAATCTCAGTGTGCGGAGCAGCATCATCTCAGTCCATCTCAGCCCTCCAGGGTTCTCC 69
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 70 TCGAAGCTTCTCGATCTTCTCGAGAGACTTGTGTCAAGCTCATG----- 114
QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 115 -----CACGATGCTTCTCTCCTCTCCAC-----TAC 141
QY 123 IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgArgAsnLeuProVal 141
Db 142 CTCTCAGGCAACTTCGCTCCCATCGGTGATGAATCTCCGTCGATCAAGGATCTCCCGTC 201
QY 142 ValGlyLysLeuProAspSerIleLysGlyValTyValArgAsnGlyAlaAsnProLeu 161
Db 202 CATGGATTTCTCCCGCAATGCTTGAATGTGAATTTGAGGGTTGGTCCAAACCCCAAG 261
QY 162 HisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValLys 181
Db 262 TTTGATGCTGCTCGCTGGATATCACTGGTTTGTGAGATGGGATGATTCATGGGTATCGC 321
QY 182 PheGluHisGlySerAlaSerTyAlaCysArgPheThrGluThrAsnArgPheValGln 201
Db 322 ATCAAGATGGGAAACCTACTTATGTTCTCGATATGTTAAGACATCAGCTCTTAAGCAG 381
QY 202 GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221

Db 382 GAAGAGTCTTCGAGAGTCCAAATTCATGAAG---ATTGGTGACCTTAAGGGGTTTTC 438
Qy 222 GlytlealaArgLeuMetLeuPheTy:AlaArgAlaAlaAlaGlyIleValAspProAla 241
Db 439 GGATTGTAAATGGTCAATGCTCAACAGCTGAGAACGAAAGCTCAAAATATTGGCAACACT 498
Qy 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTy:ProAsnGlyArgLeuLeuAlaMet 261
Db 499 TATGGAAATGGAATGCCAATACAGCACTGTATATCACCATTGGAACAACTTCTAGCATTA 558
Qy 262 SerGluAspLeuProTy:GlnValGlnIleThrProAsnGlyAspLeuLeuThrVal 281
Db 559 CAGGAGGCAGATAAGCGTACGTCAATCAAGATTGGAAGATGAGACCTGCAACACTCTT 618
Qy 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301
Db 619 GGTATAATAGATTATGACAGAGATTGACCACTCTTCACTGTCTACCCCAAAAGTTGAC 678
Qy 302 ProGluSerGlyGluLeuPheAlaLeuSerTy:AspValValSerLysProTy:LeuLys 321
Db 679 CCGTATTACGGTGAATGTTTACATTTCGGTATTTCG---CATAGCCACCTTATCTCACA 735
Qy 322 Ty:ProAsnGlyPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341
Db 736 TACAGAGTTATCTCGAAGATGGCATTATGATGACCCAGTCCCAATTTACTATATCAGAG 795
Qy 342 ProThrMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361
Db 796 CCTATCATGATGATGATTGCTATTACTGAGACTTATGCAATCTTCATGATCTTCCT 855
Qy 362 ValValPheLysLeuProGluMetIleArgGlySerProValVal---Ty:AspLys 380
Db 856 ATGCACCTTCAGGCCAAAGGAATGGTGAAGAGAGAAATGATATATATCATCTATTGATCC 915
Qy 381 AsnLysValAlaA:ArgPheGlyIleLeuAspLysTy:AlaGluAspSerSerAsnIleLys 400
Db 916 ACAAAGAGCTGTTTGGTGTCTTCAGGCTATGCCAGGATGAACCTATGATTAGA 975
Qy 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluProGlu 420
Db 976 TGGTTGAGCTTCCCACTGCTTTATTTCACAACGCCAATGCTTGGAA---GAA 1029
Qy 421 ThrAspGluValValIleGlySerCysMetThrProProAsp----- 435
Db 1030 GAGGATGAAGTCGCTCTCATCTGCTGCTGCTGAGAAATCCAGATCTTGACATGGTCAGT 1089
Qy 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455
Db 1090 GGGAAAGTGAAGAAAGAACTCGAAATTTGGCAACGAACTGTACGAAATGAGATTCAAC 1149
Qy 456 LeuLysThrGlyGluSerThrArgArgProIleLeuSerAsnGluAspGlnValAsn 475
Db 1150 ATGAAACGGGCTCAGCTCTCTCAAAAAAACTATCCGCACTGCG----- 1194
Qy 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTy:Leu 495
Db 1195 GTTGATTTCCCCAGAAATCAATGAGTGCTACACCGGAAAGAAACAGATACGTATATGGA 1254
Qy 496 AlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAsp----- 511
Db 1255 ACAATTCTGGCAGTATCGCAAGGTTTACCGGAATCATCAAGTTTGTATCTGCATGCAAG 1314
Qy 512 -----LeuThrThrGlyGluValLysLysHisLeuTy:----- 522
Db 1315 GCTGAGCAGGGAAGAAATGCTGGAAGTACGAGGTAATATCAAGGAATATATGACCTG 1374
Qy 523 GlyAsnArgTy:GlyGlyGluProLeuPheLeuProGlyGluGlyGlyGluGluAsp 542
Db 1375 GGAGAAGCAGATATGTTTCAGAGCTATCTATGTTCCGCGTGAGACGAGCAGAGAAGAC 1434
Qy 543 GluGlyTy:LeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIle 562
Db 1435 GACGCTTACTGATATCTTTGTTTCATGATGAAACACAGCGGAATATCATCGTACCTGTG 1494

Qy 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580
Db 1495 ATAGACGCAAAACAAATCGCTGCTGAACCGGTGGAGTGGTGGAGCTCCGACAGGCTC 1554
Qy 581 ProTy:GlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db 1555 CCATATGCTTCCTCCATGCTTGTGTTTACAGAGGACAACTCCAGGAACAACTCTT 1611

RESULT 10

US-09-758-269-17
; Sequence 17, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: UCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1614)
US-09-758-269-17

Alignment Scores:

Pred. No.:	2,148-97	Length:	1617
Score:	937.00	Matches:	207
Percent Similarity:	55.81%	Conservative:	105
Best Local Similarity:	37.03%	Mismatches:	199
Query Match:	29.75%	Indels:	48
DB:	10	Gaps:	13

US-09-758-269-6 (1-599) x US-09-758-269-17 (1-1617)

Qy 64 LysGlnSerSerAsnSerProAlaIleValLysProLys---AlaLysGluSerAsn 82
Db 10 AAACCTCAGTGATGGCAGCATCATCTCAGTCCATCTAGACCCCTCCAGAGGTTTCTCC 69
Qy 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 70 TCGAAGCTTCTCGATCTTCTCGAGAGACTTGTGTCAGCTCATG----- 114
Qy 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 115 -----CAGCATCTTCTCTCCCTCTCCAC-----TAC 141
Qy 123 IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgArgAsnLeuProVal 141
Db 142 CTCTCAGGCAACTCTGCTCCCATCTCGTATGATAAATCTCCCGTCAGAGATCTCCCGCTC 201
Qy 142 ValGlyLysLeuProAspSerIleLysGlyValTy:ValArgAsnGlyAlaAsnProLeu 161
Db 202 CATGGATTCTTCCCGAATGCTTGAATGTTGAGGTTGGTCCAAACCCCAAG 261
Qy 162 HisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValLys 181
Db 262 TTTGATGCTGTCGCTGGATATCATCTGTTGATGAGATGGGATGATTCATGGGTACGC 321
Qy 182 PheGluHisGlySerAlaSerTy:AlaCysArgPheThrGlnThrAsnArgPheValGln 201
Db 182 PheGluHisGlySerAlaSerTy:AlaCysArgPheThrGlnThrAsnArgPheValGln 201

Db 322 ATCAAGAGTGGAAAGCTACTTATGTTTCATATGTTAAGACATACGCTCTTAAGCAG 381
Qy 202 GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221
Db 382 GAAGAGTCTTCGGAGCTGCCAAATTCATGAG---ATTGGTGACCTTAAGGGGTTTTTC 438
Qy 222 GlyIleAlaArgLeuMetLeuPheTyraAlaArgAlaAlaGlyIleValAspProAla 241
Db 439 GGAATGCTTAATGGTCAATATCAACAGCTGAGAACGAGCTCAAAATATTGGCAACACT 498
Qy 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTyraPheAsnGlyArgLeuLeuAlaMet 261
Db 499 TATGGAATGGAACTGCCAATACACACTCGTATATACCATCGGAAACACTTCTAGCATT 558
Qy 262 SerGluAspLeuProTyroGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281
Db 559 CAGGAGGCGAGATAAGCGGTACGTCAAAAGTTTGGAGATGGAGAGCTCCCAACACTTT 618
Qy 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301
Db 619 GGTATAATAGATTATGACAGAGATTGCCACTCTTCACCTGCTCACCACCAAGTTGAC 678
Qy 302 ProGluSerGlyGluLeuPheAlaLeuSerTyraPheValValSerLysProTyroLeuLys 321
Db 679 CGGTTTACGGGTGAAATGTTTACATTCGGCTATTGCG---CATACGCCACCTTATCTCACA 735
Qy 322 TyraPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341
Db 736 TACAGATTATCTCGAAGATGGCAATTATGATGACCCAGTCCCAATTTACTATATCAGAG 795
Qy 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361
Db 796 CCTATCATGATGATTTGCTATTACTGAGACTTATGCAATCTTCATGATGATCTTCTCT 855
Qy 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValVal---TyraPheLys 380
Db 856 ATGCACCTCAGGCCAAAGAAATGGTGAAGAGAGAAATGATATATCTCAITGATCCC 915
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyraAlaGluAspSerSerAsnLys 400
Db 916 ACAAAAAGCGTCGTTTGGTGTCTTCGCGCTATGCCAGGATGAACCTTATGATTAGA 975
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu 420
Db 976 TGGTTTGGAGCTTCCCACTGCTTTATTTCCCAACAGCGCAATGCTTTGGAA-----GAA 1029
Qy 421 ThrAspGluValValIleGlySerCysMetThrProProAsp----- 435
Db 1030 GAGATGAAGTCGTCCTCATCTCTGCTGCTTGAGATCCAGATCTTGACATGTCAGT 1089
Qy 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455
Db 1090 GGGAAAGTGAAGAAACCTCGAAATTTTGGCAACGAACTGTACGAAATGAGATTCAAC 1149
Qy 456 LeuLysThrGlyGluSerThrArgProIleIleSerAsnGluAspGlnGlnValAsn 475
Db 1150 ATGAAAACGGGCTCAGCTTCTCAAAAAAACHATCCGATCTGCG----- 1194
Qy 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyraLeu 495
Db 1195 GTTGATTTCGCCAGATCAATAGTACTGCTACACCGGAAAGAAACAGAGATACGTATATGA 1254
Qy 496 AlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAsp----- 511
Db 1255 ACAATTCTGCAGAGTATCGCAAGGTTACCGGATCATCAAGTTTGAATCTGATCGATCGA 1314
Qy 512 -----LeuThrThrGlyGluValLysHisLeuTyra----- 522
Db 1315 GCTGAGACAGGGAAGAAATGCTGGAGTAGGAGGTAATATCAAGGAATATATGACCTG 1374
Qy 523 GlyAspAsnArgTyraGlyGluProLeuPheLeuProGlyGlyGlyGluGluAsp 542
Db 1375 GGAAAGGCGAGATATGCTCAGAGGCTATCTATGTTCCGCGTGAAGACACAGAGAGAC 1434

Qy 543 GluGlyTyraIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIle 562
Db 1435 GAGCGTTACTTGATATTCTTTTTCATGATGAAACACAGGGAAATCATGGTGTACTGTG 1494
Qy 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580
Db 1495 ATAGAGCBAACAACTATGCTCGCTGAACCGGTGGAGTGTGGAGCTCCGACACAGGTC 1554
Qy 581 ProTyroGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db 1555 CCATATGGCTTCATGCTTGTGTTTACAGAGGAACAACTCCAGGAACAAACTCTT 1611

RESULT 11

US-09-878-574-2543
; Sequence 2543, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael D.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2543
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-023-Q1-B1-G10
US-09-878-574-2543

Alignment Scores:
Pred. No.: 158-50 Length: 393
Score: 522.00 Matches: 95
Percent Similarity: 85.16% Conservative: 14
Best Local Similarity: 74.22% Mismatches: 19
Query Match: 16.57% Indels: 0
Gaps: 0

US-09-758-269-6 (1-599) x US-09-878-574-2543 (1-393)

Qy 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAsp 347
Db 8 GAGCAGAGGAAGTCGCGGACATAGAAATTCCTTTGACGCGCGGACGATGACGACGAC 67
Qy 348 PheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuPro 367
Db 68 TTCGATTCAGGAGAAATTTCTGGTGTATCCCGACACAGAGTGTGTTCAGCTCGGC 127
Qy 368 GluMetIleArgGlyGlySerProValValTyraPheAsnLysValAlaArgPheGly 387
Db 128 GAAATGATCAAGAGGATCGCGGTGATCTACGCGCGGAGAAATTCGCGGTTCCGA 187
Qy 388 IleLeuAspLysTyraAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCys 407
Db 188 ATATGCAAGTACGCTTCGCGGATCCAGCATCTGTGGTGGTACTCGCCACGACG 247
Qy 408 PheCysPheHisLeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIle 427
Db 248 TTCCTCTTCACCTCTTGGACCGGTGGAGGAACGCGCAAAAGGAGTGTGTGTAATA 307
Qy 428 GlySerCysMetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSer 447
Db 308 GGTTCGTCATGACACCCAGATTCATATTCATGACAGAGAGAGACTGTAAGAC 367
Qy 448 ValLeuSerGluIleArgLeuAsn 455
Db 368 GTTTTACAGAGTAAGCTGAAC 391

```
RESULT 12
US-08-976-063C-1/c
Sequence 1, Application US/08976063C
Publication No. US20020182697A1
GENERAL INFORMATION:
APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A
TITLE OF INVENTION: ACID AND THEIR USE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTRA
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976.063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 32679 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (Genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas sp.
STRAIN: HR199
FEATURE:
NAME/KEY: CDS
LOCATION: 3146..3997
OTHER INFORMATION: /gene= "ORF1"
US-08-976-063C-1
Alignment Scores:
Pred. No.: 1,128-35 Length: 32679
Score: 419.00 Matches: 160
Percent Similarity: 37.97% Conservative: 83
Best Local Similarity: 25.00% Mismatches: 229
Query Match: 13.30% Indels: 168
DB: 8 Gaps: 24
US-09-758-269-6 (1-599) x US-08-976-063C-1 (1-32679)
QY 19 HisThrGlnProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeu 38
DB 17011 CATACGGCGCCACCAAGTAGTGTTCATTCTTGAATTCCTAAACAATAGCCGCGCTCGGC 16952
QY 39 ProMetAlaSerArgValThrArgLysLeuAsnValSerAlaLeuHisThrProPro 58
DB 16951 CCATTTCAC-----ATACCACCA 16934
```

59 AlaLeuHisPheProLysGlnSerSerAsn-----SerProAlaIleValVal 74
16933 CCATTGCACCTCACACACATCTTCACTTCACCGGCACACATGAAATCCG-----GTC 16883
75 LysProLysAlaLysGluSerAsnThrLys-----GlnMetAsnLeu 88
16882 AGTCCTAAACATAACTCCACCGCGCAAGCGGCATTTAGGCCATTAGCCATTCAGAACACAAA 16823
89 PheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108
16822 GGAGACCGTGCATGCGAGATTC----- 16799
109 LysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128
16798 -----AACGGCAACGCCG-----CAATTAGTAGAGAACATTTCTC 16763
129 ProValAsnGluGlnProValArgAsnLeuProValValGlyLysLeuProAspSer 148
16762 CCCACCGGTATAGAGCGACACTTGTTCGATCTAGAGGTGACGGCGAAATCCCAAAATCA 16703
149 IleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168
16702 ATAAATGGAACGTTCTTACCGTAATACGCCAGAGCTCAAGTTACCCCAAAATATTCAC 16643
169 HisPhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188
16642 ACCTTCATAGATGGAGATGGAATGCTCTGCTTCCACTTCGAAGATGGTCATGTCGAC 16583
189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgPro 208
16582 TTCATCAGTCGCTGGGTAAACCCGCTGATTACGGCGCAAGCTAGCGGAAATTCG 16523
209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228
16522 CTATTGGCATGTACAGAAACCCCTATACCGACGACACCAAGTGTAAAGGACTA----- 16469
229 PheTyrAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248
16468 -----GACCGCACCGCTTCCCAAT 16451
249 AlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyr 268
16450 ACAAGCATCATTAGCCATCACGCGAAGGTGTGGGGTGAAGGAAGACGGCTACCGTAC 16391
269 GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGly 288
16390 -----GAATCGATCCTCGT---ACACTTGAACCTCGGACACTTCGACTTACACCGC 16340
289 GlnLeuGluSer---ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
16339 CAAGTTACCGCAAAACCCACACCGCCCATCCAAAATATGACCCCGAAACGGGTGACTTG 16280
308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327
16279 TTG-----TTCTCGGTTTCGCGAGCT 16259
328 AspGlyThrLysSerProAspVal----- 335
16258 AAGGGCAAGCAACTCCAGCATGGCCTATTACATTTGCGACACGACGCGGAAGTGACA 16199
336 ---GluIleGlnLeuAspGlnPro-----ThrMetMetHisAspPheAlaIleThrGlu 352
16198 CATGAAACTTGGTTTGAGCAGCCCTATGGCGCATTCATGCGACGACTTGGCATTAACCGA 16139
353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372
16138 AATTTGCTCATTTTCCCAATTATGCGGCGCAACACAGCGCTG---TCCCGCTCAAGGCG 16082
373 GlySerProVal---ValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLys 391
16081 AAACACCAATTATATGTGGGAGCGGAACCTGGGAGCTACATTGGCGCTACTC----- 16028

Db 1080 CTTGNAACCTCGGACACTTCGACTACGACGGCCAGTTACACGCAAAACCCACACGCC 1021
Qy HisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValSer 316
Db CATCCAAATATGACCCGAAACGGGTGACTGTGTTG----- 985
Qy 317 LysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspVal--- 335
Db 984 -----TTCTTCGGTTCGGACACTAAGGGCGAAGCAACTCCAGACATGCCC 940
Qy 336 -----GluLeuGlnLeuAspGlnPro--- 342
Db 939 TATTACATTGTCGACACGACGCAAGGTGACATGAACACTGCTGTTTGACGACCCCTAT 880
Qy 343 ---ThrMetMetHisAspPheAlaTleThrGluAsnPheValValProAspGlnGln 361
Db 879 GGCGCATTCATGACGACTTTCGCAATACCGAAATGTCCTATTTCCCAATTATGCGG 820
Qy 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProVal---ValTyrAspLys 380
Db 819 GCCACCAACAGCGCTG---TCCCGCTCAAGCGGAAACAGCAATTTATATGTGGAGCG 763
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db 762 GAATCGGCGACGATCATTTGGGTATTTTCATGTTGTAATGCTTGGGAAGTCGGAACC 649
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluLysProGlu 420
Db 708 TGGCTCAAGGACCGCGCTCGGTATTTTCATGTTGTAATGCTTGGGAAGTCGGAACC 649
Qy 421 ThrAspGluValValIleGlySerCysMetThrPro-----ProAspSer--- 436
Db 648 AAGATTATATGACCTTATGAAAGTGAATCTCGCGCTTACCGCTTCCCACTCAAA 589
Qy 437 -----Ile 437
Db 588 AACCAACCTTCGCCCTCGAGAAACCGCTACACCGCTGACTCGTTGGGAATGACCTC 529
Qy 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457
Db 528 GATGACGACGCGGACGATCAAGCGAACCCTGCTACACGATTTCTTTTCGGAATGCCA 469
Qy 458 ThrGlyGluSerThr----- 462
Db 468 ATCATGGAATTTCTGTTGCCCTCGAATGCAACCGCTATGGCTTTATGGGGTGGACGAT 409
Qy 463 ---ArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481
Db 408 CCACGCAAAACCACTTCGCGATCAGCAGCGCGGAGAGATA----- 370
Qy 482 AsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501
Db 369 -----TTCGGCTACAACTCACTCTCGGC---ATCTGG 343
Qy 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysHisLeu 521
Db 342 -----GACAAACCCAGCGAGGTGACTACGACCTCTGTGATC 310
Qy 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuPro---GlyGluGlyGlyGlu 540
Db 309 TCCGCGAAGCCTCGCGCGCCAGAGCGGCTTCGTCCTCTAGAAAGTCCGACCGCGCC 250
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db 249 GAAGGTGACGGGTACTTGTGCTGACCGTGTGTGCTCGCTCGATGAAATCGACGATCTG 190
Qy 561 GlnIleValAsnAlaValSerLeuGlu-----ValGluAlaThrValLysLeuProSer 578
Db 189 GTAATTCGACACTCAAGACATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130
Qy 579 ArgValProTyrGlyPheHisGlyThrPheIle 589
Db 129 CGGCTAAGGCGCGCTCTCCATGGCTGCTGGTA 97

RESULT 14

US-09-878-574-2872
; Sequence 2872, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2872
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(320)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-019-Q1-B1-B11
; US-09-878-574-2872

Alignment Scores:
Pred. No.: 4,59e-33 Length: 320
Score: 369.00 Matches: 68
Percent Similarity: 80.73% Conservative: 20
Best Local Similarity: 62.39% Mismatches: 15
Query Match: 11.71% Indels: 6
DB: 10 Gaps: 2

US-09-758-269-6 (1-599) x US-09-878-574-2872 (1-320)

Qy 447 SerValLeuSerGluIleArgLeuAsnLeuLysThrGlyLysThrArgArgProIle 466
Db 2 AGCGTTTAAAGAAAGAGGCTGACATGAGAACGGGAGGAGGAGGAGGAGGAGGAGG 51
Qy 467 IleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeu 486
Db 62 GTG-----GAGGAATGAACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 109
Qy 487 GlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGly 506
Db 110 GGAGAGAAACACGGTTCGCATATTTTGCATAGCGGAACCGTGGCCGAGGTGCGGG 169
Qy 507 PheAlaLysValAspLeuThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArg 526
Db 170 GTGCGAAGGTGGACCTGGAGAGTGGGAGGTGAAGAGGACGACGAGTACGAGAGAGAGG 229
Qy 527 TyrGlyGlyGluProLeuPheLeuPro-----GlyGlyGlyGlyGluAspGlyGly 544
Db 230 TTCGGTGGAGAGCATTTCTTTCACACACGCTGGAGGAATGGGAATGAGATGAAGG 289
Qy 545 TyrIleLeuCysPheValHisaspGlu 553
Db 290 TACGTGATGCTTTTGTGATGACGAG 316

RESULT 15

US-09-770-696-172/c
; Sequence 172, Application US/09770696
; Patent No. US2001004940A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang

APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krickler, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE OF INVENTION: thaliana
FILE REFERENCE: 2031US (PARA-020PRV)
CURRENT APPLICATION NUMBER: US/09/770,696
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,278
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 911
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 172
LENGTH: 200
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(200)
OTHER INFORMATION: n = A, T, C or G
US-09-770-696-172

Alignment Scores:
Pred. No.: 9,82e-31 Length: 200
Score: 346.00 Matches: 65
Percent Similarity: 98.48% Conservative: 0
Best Local Similarity: 98.48% Mismatches: 1
Query Match: 10.98% Indels: 0
DB: 9 Gaps: 0

US-09-758-269-6 (1-599) x US-09-770-696-172 (1-200)

Qy	265	AspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPhe	284
Db	199	GATTTACCTTACCAGTTCAGATCCTCCCAATGGAGATTTAAACCGTTGGTCGGTTC	140
Qy	285	AspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSer	304
Db	139	GATTTTGATGACAAATTAGAATCCCAATGATTGCCCCGCCGAAAGTCGACCCGGAATCC	80
Qy	305	GlyCluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArg	324
Db	79	GGTGAAACNNITCGGTTTAAAGTCAGCGTCGTTCAAAGCCTTACCTTAAATACTTCCGA	20
Qy	325	PheSerProAspGlyThr	330
Db	19	TTCTCACCAGCGGAACT	2

Search completed: November 17, 2003, 00:51:50
Job time : 574 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2003, 20:21:32 ; Search time 96 Seconds

(Without alignments)
2754.047 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASTATAAVSGRWLGGNHT.....VPYGFHCTFGADDLAKQVV 599

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cn2_1/usptc.spool/US09758269/runat.14112003.192309.25781/app.query.fasta.1.775

-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -XINMATCH=0.1 -LOEPCI=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09758269 @CN 1.1.56 @runat.14112003.192309.25781 -NCPU=6 -ICPU=3

-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG

-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6

-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*

1: /cn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	494.5	15.7	4403765	3	US-09-103-840A-2
2	494.5	15.7	4411529	3	US-09-103-840A-1
3	475.5	15.1	4403765	3	US-09-103-840A-2
4	475.5	15.1	4411529	3	US-09-103-840A-1
5	419	13.3	32679	4	US-08-976-063E-1
6	411.5	13.1	1518	4	US-08-976-063E-21
7	211	6.7	2629	1	US-08-200-807-1
8	211	6.7	2629	1	US-08-488-305A-1
9	210	6.7	1724	3	US-09-385-259-1
10	210	6.7	1724	4	US-09-645-370-1
11	186	5.9	36063	4	US-08-311-731A-140
12	127.5	4.0	4242	4	US-09-252-991A-7056

13	127.5	4.0	10023	4	US-09-252-991A-6997	Sequence 6997, Ap
14	123.5	3.9	5163	3	US-08-700-651-1	Sequence 1, Appl
15	123.5	3.9	5163	3	US-08-928-361B-4	Sequence 4, Appl
16	123.5	3.9	5163	4	US-09-588-995A-4	Sequence 4, Appl
17	123.5	3.9	5318	3	US-08-700-651-2	Sequence 2, Appl
18	123.5	3.9	5318	3	US-08-928-361B-3	Sequence 3, Appl
19	123.5	3.9	5318	4	US-09-588-995A-3	Sequence 3, Appl
20	116.5	3.7	5511	3	US-08-928-361B-2	Sequence 2, Appl
21	116.5	3.7	5511	4	US-09-588-995A-2	Sequence 2, Appl
22	116.5	3.7	7334	3	US-08-928-361B-1	Sequence 1, Appl
23	116.5	3.7	7334	4	US-09-588-995A-1	Sequence 1, Appl
24	115	3.7	1389	4	US-09-252-991A-11721	Sequence 11721, A
25	115	3.7	2322	4	US-09-252-991A-11519	Sequence 11519, A
26	115	3.7	47981	4	US-09-679-279-1	Sequence 1, Appl
27	110	3.5	1835	4	US-09-252-991A-2493	Sequence 2493, Ap
28	106	3.4	1839	3	US-09-461-697-76	Sequence 76, Appl
29	106	3.4	1825	3	US-09-461-697-75	Sequence 75, Appl
30	106	3.4	1953	4	US-09-252-991A-759	Sequence 759, App
31	105.5	3.3	1416	4	US-09-107-532A-2051	Sequence 2051, Ap
32	105.5	3.3	4508	5	PCT-US93-06251-34	Sequence 34, Appl
33	105	3.3	1172	3	US-08-861-774E-17	Sequence 17, Appl
34	104.5	3.3	1399	4	US-09-328-352-3002	Sequence 3002, Ap
35	104	3.3	1341	4	US-09-350-756-2	Sequence 2, Appl
36	103	3.3	1479	4	US-09-252-991A-4180	Sequence 4180, Ap
37	103	3.3	1575	4	US-09-252-991A-2009	Sequence 2009, Ap
38	103	3.3	1581	4	US-09-252-991A-654	Sequence 654, App
39	103	3.3	3102	4	US-09-252-991A-4429	Sequence 4429, Ap
40	102.5	3.3	2607	2	US-08-907-166-1	Sequence 1, Appl
41	102.5	3.3	2607	4	US-09-391-340-1	Sequence 1, Appl
42	102.5	3.3	3257	4	US-09-585-173B-39	Sequence 39, Appl
43	102.5	3.3	14672	4	US-08-961-527-111	Sequence 111, App
44	102	3.2	879	4	US-09-252-991A-12072	Sequence 12072, A
45	102	3.2	3396	4	US-09-252-991A-14676	Sequence 14676, A

ALIGNMENTS

RESULT 1

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R. M.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Alignment Scores:

Pred. No.:

Score: 4.22e-39

Best Local Similarity: 494.50

Query Match: 26.39%

DB: 15.70%

Matches: 180

Mismatches: 94

Indels: 251

Gaps: 157

US-09-758-269-6 (1-599) x US-09-103-840A-2 (1-4403765)

QY	1	MetAlaSerPheThrAlaThr---AlaAlaValSerGly-----Arg 13	
Db	751637	ATGCCACGAGGTGGCGGATCGCGGTCTGCGGACCGTCACCGCATCGAAC	751696
QY	14	TrpLeuGlyGlyAsnHisThrGlnProLeuSerSerSerGlnSerSerAspLeuSer 33	
Db	751697	TGGCGTAGCCCTCTCGGCTATGGCGCATCAGTCCGCGCATCCGCGCGAAATGGTG	751756
QY	34	TyrCysSer-----SerLeuProMetAlaSerArgValThrArgLysLeuAsnVal--- 50	
Db	751757	TACACCGCATCGTGGAGTCCCTGCTCGCGCGCCACCTTGGCGGTCTCGACGCGGTG	751816
QY	51	-----SerSerAlaLeuHisThrProAlaLeuHisPheProLysGlnSerSer 67	
Db	751817	GGCCCGTATCGTGCAGCATCGACCGCGACCGCGCGTGC-----ACGAGTCTGTG	751864
QY	68	AsnSerProAlaIleValVal-----LysProLysAlaLysGluSerAsn 82	
Db	751865	CGAACACCGGTCTCGAGGTCTATCGTCCCATGTTCTCACCAAGGGGTACCGTTCAT	751924
QY	83	ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102	
Db	751925	ATCAGTGAATACATGTTATAGAGATCGCATGACCCCGCACAGCGCGCGAA--- 751981	
QY	103	PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122	
Db	751982	-----TCCCAAAACCCATATCTCGAG	752002
QY	123	IleAlaGlyAsnPheAlaProValAsnGlnProValArgArgAsnLeuProValVal 142	
Db	752003	-----GGCTTCCTGCGCGGTGAGCAGCGAGTAAGTACCGACCGACCTCGCGTACC	752056
QY	143	GlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHis 162	
Db	752057	GGCCGATTCGGAACACCTACGACGGGGTATCTGCTAAACGGCGCCCAACCCGCGTGG	752116
QY	163	Glu-----ProValThrClyHisPhePheAspGlyAspGlyMetValHisAlaVal 180	
Db	752117	GAGGTGACCGCGGACCC---TACCACCTGGTTCACCGCGAGCGCATGTCGCGAGTC	752173
QY	181	LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200	
Db	752174	CGCGTGGCGGACGGGAAGGCC-----CGCTGTAT	752203
QY	201	GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGlu----- 216	
Db	752204	CGCAATCGTGGTGGCGACACCGCGGTGTCGCGCGCCCTCGGCGACCCATTCGCGC	752263
QY	217	---LeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAla 235	
Db	752264	CGGCTCACCGCGCACCGGATTCGAG-----	752293
QY	236	GlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsn 255	
Db	752294	-----GGCGTCCCAACACCAACGAGTGTGTCACCGCC	752326
QY	256	GlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsn 275	
Db	752327	GGACGACCTGGCTGTGGTGGCGCGGTG-----GTCACTACCACTCACC	752377
QY	276	GlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIle 295	
Db	752378	GATGAGTGGACACCGTGGACCCCTGTGACTTCGACGGCACCCCTGCACGCGCGTTACACC	752437
QY	296	AlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValVal 315	
Db	752438	GGCCATCCGCGGTGATCCGCACACGCGGTGAATGCACGCGGTGTCCTACTCTGTCGCC	752497
QY	316	SerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspVal 335	
Db	752498	CGCGACACAGAGTGCAGTACTCGGTGATCGGCACCGGACACGCTGTCGCGAGGT, 752557	
QY	336	GluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGluAsnPheVal 355	

RESULT 2
 US-09-103-840A-1
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.


```
QY 519 LysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGly 538
Db 751276 GTCGCTCGCTGATCCGACCTGCTGATCGCGAGATGGTGTGTCGCGAACCCTGCG 751335
QY 539 GlyGlu---GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys 557
Db 751336 GCGCTGAGAGATGACGGGATTCCTATGGCTACGGCTGGCACCAGCGCGGACGAA 751395
QY 558 SerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuPro 577
Db 751396 GCCAGCTGCTCTTCTGTCGATCCAGACTCTCGAGTCGATCGCCACCGTCGACCTGCCA 751455
QY 578 SerArgValProTyrGlyPheHisGlyThrPhe-----IleGly 590
Db 751456 CAGCGGTGCCGATGGCTTCCAGCGCACTGGCGCGCAGCACCTGACGGCGCTCGGG 751515
QY 591 AlaAsp 592
Db 751516 TGGCAT 751521

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: WHITE, Owen R. Robert D.
; APPLICANT: FRASER, Claire M.
; APPLICANT: FRASER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 5,52e-37 Length: 4403765
Score: 165 Matches: 165
Percent Similarity: 40.35% Conservative: 90
Best Local Similarity: 26.11% Mismatches: 255
Query Match: 15.10% Indels: 123
DB: 3 Gaps: 23

US-09-758-269-6 (1-599) x US-09-103-840A-2 (1-4403765)
QY 17 GlyAsnHisThrGlnProLeuSerSerSerClnSerSerAspLeuSerTyrCysSer 36
Db 1018941 GCGCGTACCACATCAACCCAGCGGAGGATGATCGGTGGCGGACATCCGG-----TCG 1018988
QY 37 SerLeuProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThr 56
Db 1018987 GGGCTTCGGGTGCGGATGCTGCTCGACGCGGGCAACAGGTGAGCGGATTCGCGGTG 1018988
QY 57 ProProAlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleValLysPro 76
Db 1018827 ACTATCAGGTGCGAGAACCGCGTTCGCGCAGCTGAATTCGTTGGCGGACCGGCCA 1018768
QY 77 LysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAla 96
Db 1018767 CCACGGTCAGTTTCG-----TTGTCAGCACACAC 1018738
```

Db 1017792 -----TTCTACGAAGCGGATCCGACGCCACTTGACACCGGAGGAACGAA 1017749
Qy 443 -----GluAsnLeuLysSerValLeuSerGlu1 452
Db 1017746 GTGGAGAAGCTGTTCCGTTCTCGGTCCTGGATCGCTCGACTCCGCGGTACATCCGTTG 1017689
Qy 452 eArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleLeuSerAsnGluAspG1 472
Db 1017688 GCGGCTCAATATGTTGACCGGG-----GCATCCACGAGGA 1017653
Qy 472 nGlnValAsn-----LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLy 489
Db 1017652 GCAACTGTCCGAGTCCATCACCGAGTTCGGAACCATCAACGCCGATACCGGCCACGAG 1017593
Qy 489 sThrLysPheAlaThrLeuAlaLeuAlaGluPro---TrrProLysValSerGlyPheAl 508
Db 1017592 CTACCGTTACCTATGCCGTACCGGCAACCGAGTGTTCCTGTTCCGACGACTGGT 1017533
Qy 508 aLysValAspLeuThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrG1 528
Db 1017532 TAAGCACGATCTGCTACCGGCAACCGAGTGTACTCTGTTCCGTTGACGCGCTCTACGG 1017473
Qy 528 yGlyGluProLeuPheLeuProGlyGluGlyGlu---GluAspGluGlyTyrIleLe 547
Db 1017472 AAGTGAGACCGGATGGCTCCACGGTGGGAGCAGCGCCGAGGACGACGGCTATCTGGT 1017413
Qy 547 uCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValse 567
Db 1017412 CACCCTCACCGACATGAACGACGACGATCGTATTGCTGTTTCGACGCGGCGG 1017353
Qy 567 rLeuGluValGluAlaThrValLys-----LeuProSerArgValProTyrGlyPheHi 585
Db 1017352 CCCCAGGATGGCCGATGATGACAGTGTGACGCGGACGTAATTTCCAGCGGCACGCA 1017293
Qy 585 sGlyThrPheLeuGlyAlaAspLeuAlaLys 596
Db 1017292 TTCGCGTGGGTCGCGGCGCGGAGTTCGTCGC 1017259
RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R. M.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09103,840A
; NUMBER OF SEQ ID NOS: 2
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 5,53e-37 Length: 4411529
Score: 475.50 Matches: 165
Percent Similarity: 40.35% Conservative: 90
Best Local Similarity: 26.11% Mismatches: 255
Query Match: 15.10% Indels: 123
DB: 3 Gaps: 23
US-09-758-269-6 (1-599) x US-09-103-840A-1 (1-4411529)
Qy 17 GlyAsnHisThrGlnProLeuSerSerGlnSerAspLeuSerTyrCysSer 36

Db 1018939 GCGGCTACCCATCAACCCACGCGGAGGACTGATCGTGGCGGACATCCGG-----TCG 1018986
Qy 37 SerLeuProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThr 56
Db 1018985 GGGCTTCGGTGTGGGATGCTGCTGACGCGCAACACAGTGGAGCGGATTCGCCGTG 1018826
Qy 57 ProProAlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleValValLysPro 76
Db 1018825 ACTATCAGTTCGAGAACCGCAAGCGTTCGGACGCTGAATTCGTTGGTGGCAGCGGCA 1018766
Qy 77 LysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAla 96
Db 1018765 CCACGTCAGTTTCG-----TTGTCAGCCACAC 1018736
Qy 97 LeuAspAlaAlaGluGlyPheLeuVal-SerHisGluLysLeuHisProLeuProLys-- 115
Db 1018735 GAGGATCGTGAATAGACATCAGCATTCGCGCAAGTACTTGTGACACCTTCGCGAAG 1018676
Qy 116 -ThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProVa 135
Db 1018675 GACGACCCACCTATCGCACCGGTCGTCGGGACACAGACACCGCAATGGATGCCAC 1018616
Qy 135 lArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValar 155
Db 1018615 GACCTG-ACC---ACCGTGACAGCGGAAGTCCCGCCGACCTGGAGCGCATCTACTCGG 1018560
Qy 155 gAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePheAspGlyAspG1 175
Db 1018559 CAACACCGAGAACCCGCTACACCCGCGCATTCGCGACCTACACCCCTTCGATGGCGAGG 1018500
Qy 175 yMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrG1 195
Db 1018499 CATGATCCATCGTCGCGCTCCGTCGATGGAAAGCCCTTACCGCAACCGATTATTTCG 1018440
Qy 195 nThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleG1 215
Db 1018439 CACCGATGATATCTTGGCCGAGAACGCGCGCGCGCTGTGGCGC----- 1018391
Qy 215 yGluLeuHisGlyHisThrGlyLeAlaArgLeuMetLeuPheTyrAlaArg----- 232
Db 1018390 -----GGTCTGGCAACCGGTGCACTGGCAAGCGGGAACACGG 1018350
Qy 233 ----AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyVal---AlaAsnAlaG1 250
Db 1018349 CTGGGCGCTCGTGGCTCATGAAGACGCGTCGAGCACCGACGTCATCGTCCCGAGG 1018290
Qy 250 yLeuVal-----TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLe 266
Db 1018289 TATCGCGCTGACTAGTCTTACAGTGGCGCGCATCTGATTCGATCGAC----- 1018241
Qy 266 uProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPh 286
Db 1018240 -CCGTACTCGGCCAATACGCTC-----GGCAAGGAGAGCTGGCACGGAAGGTTCCGTT 1018188
Qy 286 eAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyG1 306
Db 1018187 CGACTGGGCGTGTGCG-----GCATCTCGAAGGATAGACAAAGACCGCGCA 1018140
Qy 306 uLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSe 326
Db 1018139 ACTGTTCTTCTCACTACAGC---AAGCAAGAGCGCTATATCGCTACGCGTTCGCA 1018083
Qy 326 rProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHi 346
Db 1018082 CCAGAACAATAGACTTGTGCATCTATCTGATTCGCTGCGCGCGCGCGCTACCGCA 1018023
Qy 346 sAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLe 366
Db 1018022 TGACATGGCGTTTACCGGAAATTTACGTAATCTCTCAACGAT----- 1017983
Qy 366 uProGluMetIleArgGlyGlySerProValValTyrAspLysAsnLysVal----- 383

Db 1017982 -----TTTCCACTGTTCTGGGATCCAGGCTGCTCGAGCGCA 1017945
QY 384 -----AlaArgPheGlyIleLeuAspLysTy 392
Db 1017944 COTGACCTACACGCTCTATCCGAGATTCATCTCGGTTCGGGTGTGT 1017893
QY 392 rAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLe 412
Db 1017892 -GCTCGCGAGGCAACGACATCTGCTGTTCGAGCCGATCCAGCTTCGTGTTCACATT 1017834
QY 412 uTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetTh 432
Db 1017833 CACCAACGCTACGAG-----CAGGGGACGAGATCTGCTCGACGGC----- 1017791
QY 432 rProProAspSerIlePheAsnGluSerAsp----- 442
Db 1017790 -----TTCTACGAGGCGATCCGAGCGACCTTGACACCGGAGAACGAA 1017747
QY 443 -----GluAsnLeuLysSerValLeuSerGluI 452
Db 1017746 GTGGGAGAGCTGTTTCGGTTCCTGCTGCTGATCGGTGACGTCCCGCTACATCGGTG 1017687
QY 452 eArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspG 472
Db 1017686 GGGGCTCATATGGTCACCGG-----GCAGTCCACGAGGA 1017651
QY 472 nGlnValAsn-----LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLy 489
Db 1017650 GCAACTGTCGAGTCCATCACCAGGATTCGGAACCATCAACGCGGATTACGCGCCAGAG 1017591
QY 489 sThrLysPheAlaTyLeuAlaLeuAlaGluPro---TrpProLysValSerGlyPheAl 508
Db 1017590 CTACCGTTACACTATGCGGCTACCGGCAACCGAGTTCGTTCTGTCGACGAGTGT 1017531
QY 508 aLysValAspLeuThrThrGlyGluValLysLysHisLeuTyGlyAspAsnArgTyG 528
Db 1017530 TAAGCAGCATCTCTCACCGCAACACGAGTGTACTCTGCTGCTGACGCGCTACGG 1017471
QY 528 yGlyGluProLeuPheLeuProGlyGlyGlyGlu---GluAspGluGlyTyIleLe 547
Db 1017470 AAGTAGACCGCGATGCTCCAGGTGGGCACGACGCGCGGAGGACGCGGTATCTGT 1017411
QY 547 uCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSe 567
Db 1017410 CACCTCACCACGACATGAACGACGACGATCTATTCGCTGTTTCGACGCGCGCG 1017351
QY 567 rLeuGluValGluAlaThrValLys-----LeuProSerArgValProTyGlyPheHi 585
Db 1017350 CCGCGCGATGGCCCGATATGCAAGCTTGACCTGCGACGCGGACGATTTTCCAGCGGCA 1017291
QY 585 sGlyThrPheIleGlyAlaAspAspLeuAlaLys 596
Db 1017290 TTCGCGCTGGTGGTGGCGCGCGCGAGTTGCGTCGC 1017257

RESULT 5
US-08-976-063E-1/c
; Sequence 1, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Priefert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976,063E
; CURRENT FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1

LENGTH: 32679
TYPE: DNA
ORGANISM: Pseudomonas sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (3146) .. (3997)
OTHER INFORMATION: Gene = "ORF1"
US-08-976-063E-1
Alignment Scores:
Pred. No.: 2,58E-34 Length: 32679
Score: 419.00 Matches: 160
Percent Similarity: 37.97% Conservative: 83
Best Local Similarity: 25.00% Mismatches: 229
Query Match: 13.30% Indels: 168
DB: 4 Gaps: 24
US-09-758-269-6 (1-599) x US-08-976-063E-1 (1-32679)
QY 19 HisThrGlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrcysSerLeu 38
Db 17011 CATACGCCGCCAGATAGTGTCTTGAATTTCTAAACAATAGCCAGCTCGGC 16952
QY 39 ProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrPro 58
Db 16951 CCATTTTCA-----ATACCACCA 16934
QY 59 AlaLeuHisPheProLysGlnSerSerAsn-----SerProAlaIleValVal 74
Db 16933 CCATTGCTCTACAGACATCTTCTCACCAGCACACATGAATCG-----GTC 16883
QY 75 LysProLysAlaLysGluSerAsnThrLys-----GlnMetAsnLeu 88
Db 16882 AGTCCTAAACATAAATCCAAACCCGCGCATTTAGGCCATTAGGCCATTAGAACACAAA 16823
QY 89 PheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108
Db 16822 GGAGACGCTGCCATGCGGAGATTC----- 16799
QY 109 LysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128
Db 16798 -----AACCAGCACGACCCG-----CAATTAGTAGGAACACTTCTC 16763
QY 129 ProValAsnGluGlnProValArgAsnLeuProValValGlyLysLeuProAspSer 148
Db 16762 CCCACCGGTATAGAGGAGAGCTTGTTCATCTAGAGTTTACGCGGAAATCCCAAAATCA 16703
QY 149 IleLysGlyValTyValValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168
Db 16702 ATAAATGGAACGTTCTACCGTAAATACGCGAGGCTCAAGTTACCCACAAAAATTCAC 16643
QY 169 HisPheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188
Db 16642 ACCTTCATAGATGAGATGGAATGGCTCTGCTTCCCTTCCAGATGGTTCATGTCGAC 16583
QY 189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgPro 208
Db 16582 TTCATCAGTCTCGGTGTTAAACCGCTCGATTACGCGCGAAACGACTAGCGGAAATCG 16523
QY 209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228
Db 16522 CTATTGTCATGATACAGAAACCCCTATACCGACACACCACTGTAAAGGACTA----- 16469
QY 229 PheTyAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248
Db 16468 -----GACCGCACCGTTGCCAAT 16451
QY 249 AlaGlyLeuValTyPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTy 268
Db 16450 ACAAGCATCATTAGCATCATCGGCAAGGTGTCGCGTGAAGGAGAGCGGCTACCGTAC 16391
QY 269 GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspGly 288

Db 16390 -----GNACTGGATCTCTGCT---ACACTTGAAACTCGCGGACACATTCGACTAGCAGGC 16340

Qy 289 GlnLeuGluser---ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307

Db 16339 CAAGTTACAGCAACCCACACCGCCCATCCAAATATGACCGGAACGGGTGACTTG 16280

Qy 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327

Db 16279 TTG-----TTCITTCGGTTCCGACGCT 16259

Qy 328 AspGlyThrLysSerProAspVal-----TTCITTCGGTTCCGACGCT 16259

Db 16258 AAGGCGAAGCACTCCAGACATGGCTATTACATTGTGCACAGCAGCGCAAGGTGACA 16199

Qy 336 ---GluIleGlnLeuAspGlnPro-----ThrMetMetHisAspPheAlaIleThrGlu 352

Db 16198 CATGAACTTGGTTGACAGCCCTATGGCCATTCATGACGACTTGGCCATTACCCGA 16139

Qy 353 AsnPheValValProAspGlnValValPheLysLeuProGluMetIleArgGly 372

Db 16139 AATGGTCCATTTCCTCCCAATTATCGCGCCACCAACAGCCTG---TCCCGCCTCAAGCG 16082

Qy 373 GlySerProVal---ValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLys 391

Db 16081 AAACAGCAATTTATATGTGGAGCCGGAACCTGGCGAGCTACATTGGCGTACTC----- 16028

Qy 392 TyrAlaGluAspSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHis 411

Db 16027 CGCGCGCCAGGCGACTGCTGATTCGTGGTCAAGCAGCCGCGCTCTGGATTTCAT 15968

Qy 412 LeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMet 431

Db 15967 GTTGTGAATGCTTGGGAAGTGGCAACCAAGATTATATCGACCTTATGGAAGTGAATC 15908

Qy 432 ThrPro-----ProAspSer----- 436

Db 15907 CTGCGGTCCCTTCCCACTCAAAACCAACCTTCGCCCTTGAGAAAGCGGTACCA 15848

Qy 437 -----IlePheAsnGluSerAspGluAsnLeuLysSerVal 448

Db 15847 CGCCTGACTGTTGGGAAATGACCTCGATGACGACGACGAGATCAAGAACCCCG 15788

Qy 449 LeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThr----- 462

Db 15787 CTACAGATTCTTGGGAAATGCCAATCATGATGATTCTTGTTCGCCCTGCAATGCAAC 15728

Qy 463 -----ArgArgProIleLeuSerAsnGluAspGln 472

Db 15727 CGCTATGCTTTATGGGGTGGACGATCCACGCAACCACTTGGCGCATCAGCAGCCGAG 15668

Qy 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492

Db 15667 AAGATA-----TTC 15659

Qy 493 AlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512

Db 15658 CGGTACAACTCACTCGGC---ATCTGG-----GACAAC 15629

Qy 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532

Db 15628 CACCGAGTGACTACGACTCTGTGTTCTCCGCGCAAGCCTCGCGCGCCAGGAGCCGCGC 15569

Qy 533 PheLeuPro---GlyGluGlyGlyGluAspGlyTyrIleLeuCysPheValHis 551

Db 15568 TCGTCCCTAGAAGTCCGACCGCGCGCAAGGTGACGGGTACTGCTGACCGGTGGT 15509

Qy 552 AspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu----- 569

Db 15508 CGCTCGATGAAATCGCAGCGATCTGGTAATTCCTGCACACTCAAGACATCCAGTCTGGT 15449

Qy 570 ValGluAlaThrValLysArgValProTyrGlyPheHisGlyThrPheIle 589

Db 15448 CCCGTGGCAACCATCAAGCTCCATTCCGGCTAAGGCGCGCTCTCCATGGCTGCTGGTA 15389

RESULT 6

US-08-976-063E-21/c
; Sequence 21, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:

; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Priefest, Horst
; APPLICANT: Rabenhorst, Jürgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976,063E
; CURRENT FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: not required under old rule
; US-08-976-063E-21

Alignment Scores:

Pred. No.: 9,53e-36 Length: 1518
Score: 411.50 MatChes: 137
Percent Similarity: 40.68% Conservative: 79
Best Local Similarity: 25.80% Mismatches: 192
Query Match: 13.06% Indels: 123
DB: 4 Gaps: 19

US-09-758-269-6 (1-599) x US-08-976-063E-21 (1-1518)

Qy 118 AspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArg 137

Db 1497 GACCCG-----CAATTAGTAGGAACACTTCTCCCCCGGTATAGAGCAGACTTGTTC 1444

Qy 138 AsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGly 157

Db 1443 GATCTAGAGTTGACGGCGAAATCCCAAAATCAATAAATGGAACGCTTCTACCGTAATACG 1384

Qy 158 AlaAsnProLeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetVal 177

Db 1383 CCAGAGCCTCAAGTTACCCCAAAATTCACACCTTCATAGATGGAGATGGAATGCC 1324

Qy 178 HisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsn 197

Db 1323 TCTGCTTCCACTTCGAAGATGTCATCGACTTCATCAGTCGCTGGGTTAAACCCGCT 1264

Qy 198 ArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeu 217

Db 1263 CGATTACGCGCGCAACGACTAGCGCAAAATCGCTATTTCGCGATGTACAGAAACCCCTAT 1204

Qy 218 HisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIle 237

Db 1203 ACCGACGACACAGTGTAAAGGACTA----- 1177

Qy 238 ValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArg 257

Db 1176 -----GACCGCACCGTTGCCAATCAAGCATCATTAGCCATCAGCGCAAG 1132

Qy 258 LeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAsp 277

Db 1131 GTCGTGGCGTGAAGAGACGCGCTACCGTAC-----GAACTGGATCCTCGT---ACA 1081

Qy 278 LeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer---ThrMetIleAla 296

Db 1080 CITGAACTCGCGGACACTTCGACTACGACGCGCAAGTTACGACGCAAAACCCACACGCC 1021

Qy 297 HisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSer 316

Db 1020 CATCCAAATATGACCGGAAACGGGTGACTTGTG----- 985
Qy 317 LysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspVal--- 335
Db 984 -----TTCTTCGGTTCGGCAGCTAAGGCGGAGCAACTCAGACATGGCC 940
Qy 336 -----GluileGlnLeuAspGlnPro--- 342
Db 939 TATTACATTGTCACACGACGCGAAGGTGACATGAATCTGTTGGTGGACGCCCTAT 880
Qy 343 ---ThrMetMetHisAspPheAlaIleThrGluAsnPhaValValProAspGlnGln 361
Db 879 GCGCATTCATGACAGCTTTCGCAATACCGAAATGCTCCATTTCCCAATATATGCG 820
Qy 362 ValValPheLysLeuProGluMetileArgGlyGlySerProVal---ValTyrAspLys 380
Db 819 GCCACCAACAGCGCTG---TCCCGCTCAAGCGAAACAGCAATTTATATGTTGGAGCG 763
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db 762 GAACGGGCGACATACATTGGCGTACTC-----GCGCGCGCCGAGGCGAGTCTGATTGCG 709
Qy 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
Db 708 TGCTCAAGGACCGCGCTGCGTATTTTCATGTTGTGAATGCTTGGAAAGTCGGAACC 649
Qy 421 ThrAspGluValValIleGlySerCysMetThrPro-----ProAspSer--- 436
Db 648 AAGATTATATGACCTTATGAAAGTGAATTCCTGCCGTTCCCTTCCCACTCAAA 589
Qy 437 -----Ile 437
Db 588 AACCAACCTTCCGCTCAGAAAGCGGTACCAACCGCTATGCTTATGGGGTGGACCAT 409
Qy 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457
Db 528 GATAGCAGGCGGACGAGATCAAGCGAACCCGCTACACGATTTCTTTGCGGAAATGCCA 469
Qy 458 ThrGlyGluSerThr----- 462
Db 468 ATCATGATCTCTGTTCCCTCGAATGCAACCGCTATGCTTATGGGGTGGACCAT 409
Qy 463 ---ArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481
Db 408 CCACGCAAAACCACTTCGCGATCAGCAGCGCGGAGAGATA----- 370
Qy 482 AsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501
Db 369 -----TTCGGTACAACTCACTCGGC---ATCTGG 343
Qy 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHisLeu 521
Db 342 -----GACAACCCAGGAGTGACTACGACCTCTGGTAC 310
Qy 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuPro---GlyGluGlyGlu 540
Db 309 TCGGCGAAGCCTCGCGCGCCAGGAGCGGCTTCGTCCTCTAGAGTCCGACCGCCGCC 250
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db 249 GAAGGTGACGGGTACTTGTCTGACCGGTGTTGTCGCTCGATGATAAATCGCAGCATCTG 190
Qy 561 GlnIleValAsnAlaValSerLeuGlu-----ValGluAlaThrValLysLeuProSer 578
Db 189 GTAATTCGACACTCAAGACATCCAGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTC 130
Qy 579 ArgValProTyrGlyPheHisGlyThrPheIle 589
Db 129 CGGTAAGGCGCGCTCTCCATGCTGCTGGTA 97
RESULT 7
US-08-200-807-1
; Sequence 1, Application US/08200807

Patent No. 5573939
GENERAL INFORMATION:
APPLICANT: B Vik, Claes Olof, Eriksson, Ulf
TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
TITLE OF INVENTION: bind thereto, Nucleic Acid Sequence Coding
Patent No. 5573939
TITLE OF INVENTION: Therefor, And Uses Thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,807
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/883,539
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5573939man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 280
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2629 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
US-08-200-807-1
Alignment Scores:
Pred. No.: 5,23e-13 Length: 2629
Score: 211.00 Matches: 130
Percent Similarity: 40.03% Conservative: 99
Best Local Similarity: 22.73% Mismatches: 215
Query Match: 6.70% Indels: 131
DB: 1 Gaps: 26
US-09-758-269-6 (1-599) x US-08-200-807-1 (1-2629)
Qy 105 ValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAla 124
Db 14 ATGTCAGCAAGTGTGACATCCA-----GCT 40
Qy 125 GlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro----- 140
Db 41 GGTGTTACAAGAACTGTTGAAACTGTGGAGAACTATCTCTCAGCTCACAGCCCAT 100
Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db 101 GTTACAGGAGGATCCCTCTGCTGTAACCGGAGTCTCTCTTCATGTGGCCAGGACTC 160
Qy 161 LeuHis-----GluProValThrGlyHisPheAspGlyMetVal 177
Db 161 TTTGAGGTGGATCGGAACCATTT-----TACCACCTGTTTGTATGGCAAGCCCTCTA 214
Qy 178 HisAlaValLysPheGluHisGlySerAlaCysArgPheThrGlnThrAsn 197

Db 215 CACAAGTTTGACTTTAAAGAGGACATGTGCATACATACACAGAGGTTTCATCCGACCTGAT 274
Qy ArgPheValGlnGlu-----ArgGlnLeuGlyArgProVal 209
Db 275 GCTTACGTACGGGAATGACTGAGAAAGGATCTCATACAGAAATTTGGACCTGTGCT 334
Qy PheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPhe 229
Db 335 TTCCAGATCCCTGCAAGAAATA-----TTTCCAGGTTTTTTTCTTAC 379
Qy TyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAla 249
Db 380 TTC-----CGAGGAGTGGAGTTTACTGACAAT 406
Qy GlyLeuValTyrPhe-----AsnGlyArgLeuLeuAlaMetSerGluAspLeu 266
Db 407 GCCCTTGTTAA-TATCTACCCAGTGGGGAAGATTACTATGCTGCGACAGAGCAACTT 465
Qy ProTyr-GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPh 286
Db 466 CATTACAAGGTT-----AATCCTGAG---ACCTTGGAAACAATTAAAGCAGGTTGACCT 516
Qy eAspGlyGlnLeuGlu---SerThrMetIleAlaHisProLysValAspProGluSer-- 304
Db 517 TTGCAACTATGTCTCAGTCAATGAGGACCTGCTCACCCACATTTGAAATGATGGAC 576
Qy -----GlyLeuLeuPheAlaLeuSerTyrAspValValse 316
Db 577 TGTGTACAACATTTGTAATGCTTTGGGAAAAATTTTCAATGCTCACTAATTTGATAA 636
Qy rLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValG1 336
Db 637 GATCCACACACTACAGCAGACAGAGAGATCCA-----ATAGCAAGTCAGAGATCGT 690
Qy -----GlnProThrMetMetHisAspPheAlaIleTh 351
Db 691 TGTACAATTCCTCCCTGAGTGACCGATTCAGGCACTTACGTCCATAGTTTGGTTGAC 750
Qy rGluAsnPheValValValProAspGlnGlnValValPheLysLeuProGluMetIleAr 371
Db 751 TCCCAACTATATGTTTGTGGAGACACCGAGTCAAAATTAATCTGTTCAAGTTTCTTTC 810
Qy gGlyGlySerProValValTyrAspLysAsnLysValAlaArgPhe----- 386
Db 811 TTCATGGAGT-----CITGGGAGCCCAATTACATGGATTGTTTGAATCAATGAAC 864
Qy -----GlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTr 401
Db 865 CATGGGGGTTTGGCTTCATATTGCTGCACAAAAAAGAAAGATATATCAATAATAATA 924
Qy pileAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGluTh 421
Db 925 CAGGACCTCTCT--TTTAACTCTTTCATCATCATCACTATGAGACCATGATT 981
Qy rAspGluValValValIleGlySerCysMetThrProAspSerIlePheAsn----- 439
Db 982 T-----CTGATTGTGGATCTCTGTTGCTGGAAGGATTGAATTTGTTTATAATTATT 1035
Qy -----GluSerAspGluAsnLeuLysSerValLeuSerGluIleAr 453
Db 1036 ATATTAGCAATTTACGTGAGACATGGGAGAGGTTGAAAAA--ATGCCAGAAAGGCT 1093
Qy gLeuAsnLeuLysThrGlyGluSer-----ThrArgAr 464
Db 1094 CCTCAGCTGAATTTAGGAGATACGTACTCTCTTTGAATATTGACAAAGGCTGACACAGGC 1153
Qy gProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAs 484
Db 1154 AGAAATTTAGTCACATCCCAACACAACTGCCATCGCAATTCGTGCGAGTACGAGACC 1213
Qy nMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysVa 504
Db 1214 ATCT--GGCTGGAACCTGAGGTTCTCTTTTTCAGGCGCTCGCAAGCAATTTGAGTTTCT 1270

Qy 504 lSerGlyPheAlaLys-----ValAspLeuThr-- 513
Db 1271 CAAATCAATTACCAGAAAGTGTGGAAACCTTACACATATCATATGAGCTTGGCTTG 1330
Qy 514 -----ThrGlyGluValLysLysHisLeuTy 522
Db 1331 AATCACTTTGTTTCCAGACAGGCTCTGTAAGCTGGAACGTCAAAACCTAAAGAAACCTGGGTA 1390
Qy 522 xGlyAsp-----AsnArgTyrGlyGlyGluProLeuPheLeuPro--GlyGluGlyG1 539
Db 1391 TGGCAA-GAGCCTGATTATACCTCCCTCAGACCTATCTTTGTTCTCACCAGATGCCTT 1449
Qy yGluGluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys----- 557
Db 1450 CGAGGAAGATGACGGTGTAGTTCTGAGTGTGTGTGAGCCCTGGGGCAGGACAAAGGCC 1509
Qy 558 -SerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuPr 577
Db 1510 TGTCTATCTTCGATCTGAAITGCCAAGACTTGAGTGAAGTTGCCAGGGCTGAAGTGA 1569
Qy 577 oSerArgValProTyrGlyPheHisGlyThrPhe 588
Db 1570 GATTAACTATCCCGTCACCTTTTCATGGACTGTTT 1603

RESULT 8
US-08-488-305A-1
; Sequence 1, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof; Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vineet
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; US-08-488-305A-1

Alignment Scores: 5,23e-13 Length: 2629
Pred. No.: 211.00 Matches: 130
Score:

Percent Similarity: 40.03% Conservative: 99
 Best Local Similarity: 22.73% Mismatches: 215
 Query Match: 6.70% Indels: 131
 DB: 1 Gaps: 26

US-09-758-269-6 (1-599) x US-08-488-305A-1 (1-2629)

QY 105 ValSerHisGluLysLeuHisGluProLeuProLysThrAlaAspProSerValGlnIleAla 124
 DB 14 ARGICACCCAGGTGAACATCCA-----GCT 40

QY 125 GlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro-----140
 DB 41 GGTGTTTACAAGAAATGTTTGAACCTGTGGAGGAATATCCTCACCGCTCACAGCCCAT 100

QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
 DB 101 GTTACAGCAGGATCCCTCTGCTACCGGAGTCTCTCTGATGTGGCCAGGACTC 160

QY 161 LeuHis-----GluProValThrGlyHisPhePheAspGlyAspGlyMetVal 177
 DB 161 TTTGAGGTTGGATCGGAACCATTT-----TACCACCTGTTGATGGCAAGCCCTCCTA 214

QY 178 HisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsn 197
 DB 215 CACAAGTTTGACTTTAAGAAGGACATGTACATACCACAGAGGTTTATCCGACTGAT 274

QY 198 ArgPheValGlnGlu-----ArgGlnLeuGlyArgProVal 209
 DB 275 GCTTACGTCAGCGCAATGACTGAGAAAGGATGCTCATACAGAAATTTGGCCTGTGCT 334

QY 210 PheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPhe 229
 DB 335 TTCCAGATCCCTGCAAGAATA-----TTTCCAGGTTTTTTCTTAC 379

QY 230 TyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAla 249
 DB 380 TTC-----CCAGAGTGGAGGTTACTGCAAT 406

QY 250 GlyLeuValTyrPhe-----AsnGlyArgLeuLeuAlaMetSerGluAspAspLeu 266
 DB 407 GCCCTTGTTAA-TATCTACCCAGTGGGGAAGATTACTATGCTGCGACAGACCACTT 465

QY 267 ProTyr-GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspH 286
 DB 466 CATTACAAGGTT-----AATCTGAG--ACCTTGGAAACAAATTAAAGCAGGTTGACCT 516

QY 286 eAspGlyGlnLeuGlu-----SerThrMetIleAlaHisProLysValAspProGluSer-- 304
 DB 517 TTGCAACTATGCTCAGTCAATGGAGCCACTGCTCACCCACATTTGAAATATGAGGAC 576

QY 305 -----GlyLeuPheAlaLeuSerTyrAspValValse 316
 DB 577 TGTTCACAACTGGTAATGCTTTGGAAATAATTTTCAATTGCTTACAAATTTGATAA 636

QY 316 rLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValG1 336
 DB 637 GATCCACCACTACACAGACAGCAAGGAAGATCCA-----ATAAGCAAGTCAGATCGT 690

QY 336 uileGlnLeuAsp-----GlnProThrMetMetHisaspPheAlaIleH 351
 DB 691 TGTACAATTCCTCGATGACCGGATTCAGCCATCTTACGTCATAGTTTGGTTGAC 750

QY 351 rGluAsnPheValValProAspGlnGlnValPheLysLeuProGluMetIleAr 371
 DB 751 TCCCAACTATATGTTGTTTGGAGACACCAAGTCAAAATTAATCTGTTCAAGTTTCTTC 810

QY 371 gGlyGlySerProValValTyrAspLysAsnLysValAlaArgPhe-----386
 DB 811 TTCATGGAGT-----CTTTGGGAGCCAAATTAATGATGTTTGAATCCCAATGAAC 864

QY 387 -----GlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTr 401

DB 865 CATGGGGTTTGGCTTCATATATCTGACAAAAAAGAGTATATCAATAATAATA 924
 QY 401 pIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGluTh 421
 DB 925 CAGGACCTCTCCT---TTTAACTCTTTTATCATCATCAATACCTATGAAGACCATGAGTT 981

QY 421 rAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsn-----439
 DB 982 T-----CTGATTTGGAATCTGTTCTGGAAAGGATTTGAATTTGTTTATATATTT 1035

QY 440 -----GluSerAspGluAsnLeuLysSerValLeuSerGluIleAr 453
 DB 1036 ATATTTAGCCAATTTACGTGAGAACTGGGAAGAGGTGAATAAAA---ATCCAGAAAGGCT 1093

QY 453 gLeuAsnLeuLysThrGlyGluSer-----ThrArgAr 464
 DB 1094 CCTCAGCTGAGATTAGGATAGTACTCTCTTGAATATTGACAGGCTGCACAGC 1153

QY 464 gProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAs 484
 DB 1154 AAGAATTTAGTCACTCCCAACACAACTGCCAATCTGTGCAATCTGTGCAAGTGCAGAGAC 1213

QY 484 nMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTyrProLysVa 504
 DB 1214 ATCT--GGTGGAACTGAGGTTCTTTTCAGGCTTCGCAAGCATTTGAGTTTCT 1270

QY 504 lSerGlyPheAlaLys-----ValAspLeuThr-- 513
 DB 1271 CAAATCAATTACCAGAAATGATGTGGAAACCTTACACATATGCATATGACTTGGCTTG 1330

QY 514 -----ThrGlyGluValLysLysHisLeuTy 522
 DB 1331 AATCAGCTTTGTCAGACAGGCTCTCTAAGCTGAACGCTCAAACTAAAGAAACCTGGGTA 1390

QY 522 rGlyAsp-----AsnArgTyrGlyGlyGluProLeuPheLeuPro--GlyGluGlyGl 539
 DB 1391 TGGCAA-GAGCCTGATTACATCCCTCAGAACTATCTTTGTTCTTCAACCCAGATGCCTT 1449

QY 539 yGluGluAspGlyGlyTyrIleLeuLysPheValHisAspGlyLysThrTrpLys-----557
 DB 1450 GGAGGAAGATGACGGTGTAGTCTGAGTGTGGTGGTGGCCCTGGGGCAGGACAAAGCC 1509

QY 558 -SerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuPr 577
 DB 1510 TGCTTATCTTCTGATTTGAAATGCCAAGACTTTGAGTGAAGTTGCCAGGGCTGAAGTGA 1569

QY 577 oSerArgValProTyrGlyPheHisGlyThrPhe 588
 DB 1570 GATTAACATCCCCGCTCACCTTTTCATGCACTGTTTC 1603

RESULT 9

US-09-385-259-1
 ; Sequence 1, Application US/09385259
 ; Patent No. 6201114
 ; GENERAL INFORMATION:
 ; APPLICANT: Aguirre, Gustavo D.
 ; APPLICANT: Acland, Gregory M.
 ; APPLICANT: Ray, Kunal
 ; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
 ; TITLE OF INVENTION: IN DOGS
 ; FILE REFERENCE: 19603/2481
 ; CURRENT APPLICATION NUMBER: US/09/385,259
 ; CURRENT FILING DATE: 1999-08-30
 ; EARLIER APPLICATION NUMBER: 60/103,219
 ; EARLIER FILING DATE: 1998-10-06
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1 1724
 ; TYPE: DNA
 ; ORGANISM: Canis familiaris
 ; US-09-385-259-1

Alignment Scores: 3.3e-13 Length: 1724
Pred. No.: 210.00 Matches: 113
Score: 37.08% Conservatives: 98
Percent Similarity: 37.08% Mismatches: 218
Best Local Similarity: 19.86% Indels: 140
Query Match: 6.67% Gaps: 25
DB: 3

US-09-758-269-6 (1-599) x US-09-385-259-1 (1-1724)

QY 120 SerValGlnIle-----AlaGlyAsnPheAlaProValAsnGluGlnProValArg 136
DB 31 TCCATCCCAAGTGGAGCATCCCGCGCGGTACAGAAAGCTGTTTGAACCGTGGAAAGAG 90
QY 137 ArgAsnLeuPro-----ValValGlyLysLeuProAspSerIleLysGlyVal 152
DB 91 CTGTCGTCCGCGCTCACCGCCACGTCAGAGGAGTCCGCTCTGCTCACGGCAGT 150
QY 153 TyrValArgAsnGlyAlaAsnProLeuHis-----GluProValThrGlyHisHis 169
DB 151 CTCCTCCGATCGGACCGCGGCTCTTCGAGTGGATCTGAACCATTT-----TACCAC 204
QY 170 PhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyr 189
DB 205 CTGTTTGACGACAGCCCTTCGACAAAGTTCGACTTTAAAGAAAGACACGTCACCTAT 264
QY 190 AlaCysArgPheThrGlnThrAsnArgPheValGlnGlu----- 202
DB 265 CACAGAAGGTTTCATCCGACCGATGCTTACGTCGGGCAATGACCGAGAAAGATCGTC 324
QY 203 ---ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221
DB 325 ATAACGGAATTTGGACCTGTCGTTCCAGATCCCTGCAGAAATATA----- 372
QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAla 241
DB 373 ---TTTCCAGGTTTTTTCTTACTTC----- 396
QY 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAlaMet 261
DB 397 CGAGGAGTGGAGTCACTGACAAATGCCCTTGTT-----AAGCTTACCAGTA 444
QY 262 SerGluAsp-----AspLeuProTyrGlnValGlnIleThrProAsnGly 276
DB 445 GGGGAAGATTACTACGCTCGCAGGAGACCAACTTCATTACAAAGATTAACTCTGAG--- 501
QY 277 AspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu---SerThrMetIle 295
DB 502 ACCCTGACACAAATTAAAGCAGGTGTGCTCGCAACTACGTCCTCTGTCATGGAGCCACC 561
QY 296 AlaHisProLysValAspProGluSer-----GlyGlu 306
DB 562 GCTCACCCCCACATTGAAATGATGGGACTGTGTTACACATTGGTAATGCTTTGGAAA 621
QY 307 LeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSer 326
DB 622 AATTTTTCGATTCCTACAAATATTGTAAGATCCCTCCACTCCAAAGCAGACAAAGAAAT 681
QY 327 ProAspGlyThrLysSerProAspValGluIleGlnLeuAsp-----Gln 341
DB 682 CCA-----ATAGCAAGTCCGAGTTCGTCGACATTCCTCCGAGCAGCCGATTCAAG 735
QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361
DB 736 CCATCGTAGCTCCATAGTTTGGTTTGGTTCCTCCCACTATATTGTTTGTGGAGACGCCA 795
QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySer----- 374
DB 796 GTCAAAATTAACCTGCTCAAGTTCCTTTCTCGTGGAGTCTTTGGGAGCCCACTACATG 855
QY 375 -----ProValValTyrAspLysAsnIysValAla 384
DB 856 GATTGTTTTGAGTCCAAATGAACCATCGGGGTTGGCTTCACATCGCTGACAAAAAAGA 915

385 ArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTyrPheAspAla 404
DB 916 AAAAAGTATCTCAATAAAGTACAGGACCTCTCTCTTAATCTC----- 960
QY 405 ProAspCysPheCysPheHisLeuThrPheAsnAlaTyrGluGluProGluThrAspGluVal 424
DB 961 -----TTCCATCATATCATATCTTACGAAAGACAATGAGTTT-----CTG 999
QY 425 ValValIleGlySerCysMetThrProProAspSerIlePheAsnGluSerAsp---Glu 443
DB 1000 ATTGTGGATCTCTGCTCTGGAAGGATTGAATTCGTCTACAAATTACTTGTATTATGACC 1059
QY 444 AsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArg 463
DB 1060 AATTTTACGTGAGAACTGGGAAGAGGTGAAA-----AAAAATCCCAAGAAAGGCTCCG 1110
QY 464 ArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArg 483
DB 1111 CAGCCTGAAGTAGGAGATCCGTGCTTCCTTTGAATATATCGACAAGCCGACACAGCAAG 1170
QY 484 AsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu----- 497
DB 1171 AACCTAGTCACTCTCCCAACACGAGCGCCACTGCAACTCTGGCAGCCAGACCATC 1230
QY 498 ---AlaGluProTyrProLysValSerGly-----PheAlaLysVal 510
DB 1231 TGCTCGAACCTGAGTCTCTCTTCAGGCGCTCGTCAAGCCTTTGAGTTTCTCAAAATC 1290
QY 511 AspLeuThrThr-----GlyGluValLysLysHisLeuTyrGly----- 523
DB 1291 AACTATCAGAAGTATGGCGGAAGCCTTACAGTACGCGTATGGACTTGGCTTGATATAC 1350
QY 523 ----- 523
DB 1351 TTGTTCCGAGACAGGCTCTGCAAGCTGAAGCTCAAGCTAAAGAAACGTTGGTATGCAA 1410
QY 524 ---AspAsnArgTyrGlyGlyGluProLeuPheLeuPro---GlyGluGlyGlyGluGlu 541
DB 1411 GAGCCGCACTCATACCATCAGAACCATCTTTGTTTCTCACCAGATGCTTTGGAAGAA 1470
QY 542 AspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys-----SerGlu 559
DB 1471 GATGATGTTGATGTTCTGAGTGTGGTGGAGCCCTGGGCGAGCAAAAGCTGCTTAT 1530
QY 560 LeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArg 579
DB 1531 CTCTGATTTCTGAATGCCAAGGATTGAGTGAAGTTGCCAGGCTGAAGTGGAGATTAAAC 1590
QY 580 ValProTyrGlyPheHisGlyThrPhe 598
DB 1591 ATCCCTGTCACCTTTCATGAGCTGTTTC 1617

RESULT 10
US-09-645-370-1
; Sequence 1, Application US/09645370
; Patent No. 6428958
; GENERAL INFORMATION:
; APPLICANT: Aguirre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; APPLICANT: Ray, Kunal
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; FILE OF INVENTION: IN DOGS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/645,370
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/385,259
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: 60/103,219
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1

```

; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Canis familiaris
US-09-645-370-1

```

Alignment Scores:		
Pred. No.:	3.3e-13	Length: 1724
Score:	210.00	Matches: 113
Percent Similarity:	37.08%	Conservative: 218
Best Local Similarity:	19.86%	Mismatches: 98
Query Match:	6.67%	Indels: 140
DB:	4	Gaps: 25

US-09-758-269-6 (1-599) x US-09-645-370-1 (1-1724)

Qy	120	SerValGlnIle-----AlaGlyAsnPheAlaProValAsnGluGlnProValArg	136
Db	31	TCACATCAAGATGGAGCATCCCGCGGCGGTTCACAGAAGCTGTTTGAAACCGTGAAAGAG	90
Qy	137	ArgAsnLeuPro-----ValValGlyLysLeuProAspSerIleLysGlyVal	152
Db	91	CTGTGCTGCCCGCTCACGCCCCAGCTGACAGGACAGGATCCCGCTCTGGCTCAGGGCAGT	150
Qy	153	TyrValArgAsnGlyAlaAsnProLeuHis-----GluProValThrGlyHisHis	169
Db	151	CTCCTCCGATGGCGACCGGGGCTCTTCAGAGTTGGATCTGAAACCATTT-----TACCAC	204
Qy	170	PhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyr	189
Db	205	CTGTTTGACGGACAGCCCTTCTGCACAACTTCGACTTTAAAGAGGACACGTCACCTAT	264
Qy	190	AlaCysArgPheThrGlnThrAsnArgPheValGlnGlu-----	202
Db	265	CACAGAAGGTTCATCCGACCCGATGCTTACGTCGGGCAATACCGAGAAAGGATCGTC	324
Qy	203	--ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr	221
Db	325	ATACCGAATTGGCACCTGTGCGTTCCAGATCCCTGCAGAATATA-----	372
Qy	222	GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspProAla	241
Db	373	---TTTTCACAGTGTTCCTTACTTC-----	396
Qy	242	HisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMet	261
Db	397	CGAGGATGGAGTCTACTGACAAATGCCCTTGT-----AACGTCATCCCCAGTA	444
Qy	262	SerGluAsp-----AspLeuProTyrGlnValGlnIleThrProAsnGly	276
Db	445	GGGGAAGATTACTACGCTGCGACGGAGACCACTTCATTACAAAATTAATCTCTGAG--	501
Qy	277	AspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu---SerThrMetIle	295
Db	502	ACCTGGAGACATTTAAGCAGGTTGATCTCTGCACTACGTCCTGTCTCAATGGAGCCACC	561
Qy	296	AlaHisProLysValAspProGluSer-----GlyGlu	306
Db	562	GCTCACCCCCACATTGAAATGATGAGCATGTTTACACATTGGTAAATGTGTTTGGGAA	621
Qy	307	LeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSer	326
Db	622	AATTTTTCGATTGCTTCACAAATATGTAAAGATCCCTCCACTCCAGCAGACAGGAAGAT	681
Qy	327	ProAspGlyThrLysSerProAspValGluIleGlnLeuAsp-----Gln	341
Db	682	CCA-----ATAAGCAAGTCCGAGGTGCGTACATAATTCCTCCGAGCGACCGATTCAAG	735
Qy	342	ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln	361
Db	736	CCATCGTACGTCACATGTTTGTGTTTGATCCCACTATATTGTTTGTTCGAGACGCCA	795
Qy	362	ValValPheLysLeuProGluMetIleArgGlyGlySer-----	374

STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311.731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 140:
SEQUENCE CHARACTERISTICS:
LENGTH: 36063 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-140

Alignment Scores:
Pred. No.: 2,74e-08 Length: 36063
Score: 186.00 Matches: 133
Percent Similarity: 38.61% Conservative: 72
Best Local Similarity: 25.05% Mismatches: 209
Query Match: 5.90% Indels: 121
Gaps: 23
DB:

US-09-758-269-6 (1-599) x US-08-311-731A-140 (1-36063)

QY 110 LeuHisProLeuProLysThrAlaAspProSerValGlnileAlaGlyAsnPhaAlaPro 129
DB 22576 TTGTCAGACATATAGAGGGATCAGACATGATGTGCAATTTGTCAGCAATGCTCTCT 22635
QY 130 Val-----AsnGlnProValArgAsnLeuPro----- 140
DB 22636 ACCCTGTCCGAGACGACGACGACCTCGCGCGGATTTGTCGCGGACCGACGACGATC 22695
QY 141 -----ValValGlyLysLeuProAspSerIleIysGly 151
DB 22696 GAATGGGACGCCAAGCAGCTCAACGCTTTAGCGGAGAAATACCCAGCTACCTAAAGCG 22755
QY 152 ValTyValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePhe 171
DB 22756 TACCTACCGACGCAACACTAAACCACTGACCGCGGATTCACAGTGTATACCCCGTTC 22815
QY 172 AspGlyAspGlyMetValHisAlaValIysPheGluHisGlySerAlaSerTyAlaCys 191
DB 22816 GATGAAGACGGCATGCTGCATGCTCGGCTCCGCGATGAAAGCCCTTTTTCGCAAC 22875
QY 192 ArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhePro 211
DB 22876 AGCTTTATACGACAGACGGATTTTGTCCGAGACATCGCAGGCGGCGCTGTGGCCC 22935
QY 212 LysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyAla 231
DB 22936 GG-TTG-----GGCAGAACCGGT----- 22952
QY 232 ArgAlaAlaAlaGlyIleValAspProAla---HisGlyThrGlyValAlaAsnAlaGly 250
DB 22953 -----GCATATATGACCAAGCGTGAACAGGCGCTGGGATCCGCAACGAGATG 23000

QY 251 LeuValTyPheAsn-----GlyArgLeuLeuAlaMetSerGlu----- 263
DB 23001 AAGGACGGGTTGAACACCGACGCTCAGCTTCATCGAGGATCGCGCTACCGAGTTTCTAC 23060
QY 264 -----AspAspLeuProTyGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281
DB 23061 CATTGACGAGTCTG--TACCACATAGAACCTATCCC-----GCCAATACGCGA 23108
QY 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThr---MetIle-AlaHisProLysVa 300
DB 23109 GGCAAGGAGATCGGGGACGGCGTTTACAGTTGATGGCGGTGGCGGACATCTCAAACT 23168
QY 300 LaspProGluSerGlyGluLeuPheAlaLeuSerTyArgValValSer-LysProTyL 320
DB 23169 GGCAACAGGACCGGAGACTG-----CTGTTCTTTTAAATACACAGCAGGACCGGTACA 23222
QY 320 euLysTyPheArgPheSerProAspGlyThrLys---SerProAspValGluIleGlnL 339
DB 23223 CATCGGCTATGCCCTCAACCAACCAAGCAACGAACTGACACACACAGTCGACATTCGCG 23282
QY 339 euAspGlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProA 359
DB 23283 TCGTTGGTTCGGGTTGGCAGATGCGATGGCGGTTTCCCGAAACTACAGATCTCAATA 23342
QY 359 spGlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyA 379
DB 23343 ATTG-CCCATTGTTCTGAAG-----TCCAAGACTGCTCAAG 23377
QY 379 spLysAsnLysValAlaArgPhe-----GlyIleLeuAspLysTyArgAlaGluAspSerS 397
DB 23378 AACATGTGTGCTACACGATTTCTATCGTGACATGCTCTCGCTTCGGGTGTT-TCNA 23436
QY 397 er-----AsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnA 415
DB 23437 GCACCGGCGACATCGGTGATTCGAGGGGGAATATATTTTACCGCATTTTCTCAACG 23496
QY 415 laTrpGluGluProGluThrAspGluValValIleGlySerCysMetThrProProA 435
DB 23497 CCTACGAA-----AAGGTGACGAGATCATG-----CGCG 23526
QY 435 spSerIlePheAsnGluSerAspGluAsnLeu-----LysSerValL 449
DB 23527 ATAGATTTCTGAAGACTAGCCACACACCTTTGCACGCGGCGCCTAGCAAGCCGG-C 23585
QY 449 euSerGluIleArgLeuAsnLeuLysThrGlyLysThrArgArgProIleIleSera 469
DB 23586 TGCACAGCTGGCGATTCACCTTAGTCACCGGTGCTGTAAACAGAGTAAGTACG----- 23638
QY 469 snGluAspGlnGlnValAsnLeuLeuAlaGlyMetValAsnArgAsnMetLeuGlyArgL 489
DB 23639 -----CCGGAATCTATCACCGAGTTCGGGATGATCAATGCTGATGCTCCAGCA 23690
QY 489 yThrLysPheAlaTyLeuAlaLeuAlaGluProTyPProLysValSerGly----- 506
DB 23691 AGTATCACTACTCTATACCTATGCGGCC-----ACCGGCAATCGGCTAGTTCT 23741
QY 507 -----PheAlaLysValAspLeuThrThrGlyGluVal-LysLysHisLeuTy 522
DB 23742 TCTTACAGCTATTGGCTAAACACGTTCTGCTACCAACACCTAGAGGCGCTACTCTTC 23801
QY 523 GlyAspAsnArgTyGlyGlyGluProLeuPheLeuProGlyGlyGlyGlu---Glu 541
DB 23802 GGCATGTCATCTACGGAAGTGAAGGCGGCTGGGCTCCGCGGTGAGTAGTACCGGTGAA 23861
QY 542 AspGluGlyTyIleLeuCysPheValHisAspGluLysThrTyPlys----- 557
DB 23862 GATGACGGCTAACTGGTAAACCCCTACCCAGCATGAATGCCGATAGTCTTAATTATAT 23921
QY 558 -----SerGluLeuGlnIleValAsn 564
DB 23922 GGGCTTTCGAAGCAGCCAGGATCACCAGACAGTCCAGGTGAGTAAACTTCACTGCCGAAA 23981

```

Qy 565 AlaValSerLeuGluValGluAlaThr 573
Db 23982 CGTATTTCCAGCAGCGCGCTTCCACC 24008

RESULT 12
US-09-252-991A-7056/c
; Sequence 7056, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7056
; LENGTH: 4242
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-7056

Alignment Scores:
Pred. No.: 0.00236 Length: 4242
Score: 127.50 Matches: 104
Percent Similarity: 35.96% Conservative: 65
Best Local Similarity: 22.13% Mismatches: 174
Query Match: 4.05% Indels: 129
DB: 4 Gaps: 21

US-09-758-269-6 (1-599) x US-09-252-991A-7056 (1-4242)
Qy 23 ProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSer 42
Db 2486 CATGCACCTGCGCGGTCATCGAGTCTTCGGCATGC-----GCGCGC 2442
Qy 43 ArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeuHisPhe 62
Db 2441 AGGATTGCGAACTGCACCTTCTTATTCGATCGACTTCGACGCTGCCAGCGCGCTGGTGG 2382
Qy 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGlnSerAsn 82
Db 2381 CCCCGTTGTGTGCG--CGGCCCGGTGCTGCTCGGCGCCAGGCCAGTGGCGCGCG 2325
Qy 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 2324 AGGAAA-----TCTGCAG-CTGATCCGCGCGGAGGCG 2293
Qy 103 PheLeuValSerHisGlnLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 2292 -----GTGAG-CATCTCGGCTTACCCCGAGCTACGG----- 2261
Qy 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142
Db 2260 -----CAGCCAGCTCGC----- 2240
Qy 143 GlyLysLeuProAspSerIleLysGlyValTyrValArgAsn-GlyAlaAsnProLeuHi 162
Db 2239 GGAGAGCCAGGCGCGGAGTTCGGTGGCATGTCATCCACCGCGCGC----- 2191
Qy 162 sGluProValThrGlyHis----- 2170
Db 2190 -GAAGCGCTACCGCGGAGTTCGGTGGCATGTCATCCACCGCGCGC----- 2132
Qy 170 sPheAspGlyAspGlyMetValHisAlaVal----- 180
Db 2131 CTTCAATGCTACGCGCCAGCCAGACGGTGGTGCATCGCGCTGCGCTGCGTGGCTCCGGA 2072
Qy 181 -LysPheGluHisGlySerAlaSerTyrAlaCysA-gPheThrGlnThrAsnArgPheVa 200

```

```

Db 2071 ACCACTGGAGGAGGTGCGCGCAGCGTCCGATCGGACGCGTGTGCGCGCGCGGTGCG 2012
Qy 200 lGlnGluArgGlnLeuGlyArgProValPheProLys---AlaIleGlyGluLeuHis-- 218
Db 2011 CTACATCTCGATGCGGACCTGGCCCTGGTCCGCGAGCGCGACCGCGCAACTCTAGT 1952
Qy 219 -GlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGly---- 236
Db 1951 CGGCGCGCGCGCTGGCGCGC-----GGTACCATGAGCGTCCGCGCGCTCAGCGCGCA 1898
Qy 237 -----IleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPh 254
Db 1897 GCGCTTCGTCGCGCATCTTC-----GCTCCGAGGCGCGCGCTGTACCG 1850
Qy 254 eAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnValGlnIleThrPr 274
Db 1849 CACCGCGGACCTG-----GTGCGCTGTGGA 1823
Qy 274 oAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp----- 287
Db 1822 CAACGCGCAGGTGGAATATGTCGCGCGCATCGACCAAGGAGATCCGTGGCTTCG 1763
Qy 288 -----GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSe 304
Db 1762 CATCGAACTGGCGGAGATCGAGCGCGCTGCTGGAGCATCCGAGGT----- 1714
Qy 304 rGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheAr 324
Db 1713 -CGCGAAGCGCTGCTGCGCTCGACAGCCGCGGAGCGAGTGGCGGTATGT 1655
Qy 324 gPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMe 344
Db 1654 CGCCAGCGGTGGCGGAGCAGGAGCAGCGCGCGCTGCGCGAGCGCTGAA 1595
Qy 344 tMetHisAspPheAlaIleThrGluAsnPheValValValProAspGlnGlnValPh 364
Db 1594 AACGCTATCTCAAGACGAATGCGGACATACATGTCGCGCGCGCCACTCTGTGCTCG 1535
Qy 364 eLysLeuProGluMetIleArgGlyLysSerProValValTyrAspLysAsnLysValal 384
Db 1534 CAGCTTGGC-----CTGAC 1520
Qy 384 aArgPheGlyIleLeuAspLysTyrAla-----GluAspSerSerAsnIleLy 400
Db 1519 CGCAACGCGCAAGCTCGACCGCGCGCTGCGCGCGCGCGCGCGCTCAACCGGCA 1460
Qy 400 sTriPleAspAlaPro---AspCysPheCysPheHisLeuTyrAsnAlaTrrpGluGluPr 419
Db 1459 GGCTACGAGCGCGCGCGAGCGTGTGGAGCAGCAACTGGCGGCGGTCTGGCGCGAGGT 1400
Qy 419 oGluThrAspGluValValValIleGlySerCysMetThrProProAspSerIlePheAs 439
Db 1399 GCTGACGTCGAGCGGTAGGTCTCGC-----GACAACTTCTTCGA 1358
Qy 439 nGluSerAspGluAsnLeuLysSerVal 448
Db 1357 ACTGGCGCGCATTCGATCTGTCGATC 1330

RESULT 13
US-09-252-991A-6997
; Sequence 6997, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6997
; LENGTH: 10023
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6997

Alignment Scores:
Pred. No.: 0.0102 Length: 10023
Score: 127.50 Matches: 104
Percent Similarity: 35.96% Conservative: 65
Best Local Similarity: 22.13% Mismatches: 174
Query Match: 4.05% Indels: 129
DB: 21

US-09-758-269-6 (1-599) x US-09-252-991A-6997 (1-10023)

QY 23 ProLeuSerSerSerGlnSerSerAspLeuSerTyrcysSerSerLeuProMetAlaSer 42
Db 6671 CCATGCACTGCGCGGGTATCGATCGCTTCGCGCATGC-----GCGCCG 6715
QY 43 ArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHisPhe 62
Db 6716 AGGATTGCGAACTGCACCTTCTATTTCGATCAACTTCGACGCTGCCAGCGAGCGCTGCTGG 6775
QY 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGluSerAsn 82
Db 6776 CCCGTTCTCTGCG---GCGCCGCGTGTGCTGCGGCCGCCAGCGAGTGGGCGCG 6832
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 6833 AGGAAA-----TCTGCGAG-CTGATCCGCGCGCGAGGCG 6864
QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 6865 -----GTGAG-CATCCTCGGCTTACCCCGAGTACGG----- 6896
QY 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142
Db 6897 -----CAGCCAGCTCGC-----CCAGTGGCT 6917
QY 143 GlyLysLeuProAspSerIleLysGlyValTyValArgAsn-GlyAlaAsnProLeuHi 162
Db 6918 GGAGAGCCAGGCGCGCGAGTTCGCGGTGCGATGTCATCACCAGCGCGC----- 6966
QY 162 sGluProValThrGlyHisHis-----ph 170
Db 6967 -GAAGCGTGCACGCGCGAGCACTCAACGGATTCCGACGCGCTTCGCCCGCGCTGCTT 7025
QY 170 epheaspGlyaspGlyMetValHisAlaVal----- 180
Db 7026 CTTCAATGCTACGAGCCAGCAGCAGCGTGGTCAATCGCTGCGCTGCTGCTGCTGCTGCTG 7085
QY 181 -LysPheGluHisGlySerAlaSerTyzAlaCysArgPheThrGlnThrAsnArgPheVa 200
Db 7086 AGCATGTGAGGAGGTGCGCGAGCGTCCGATCGCAGCGTGTGCGCGCGCGGGTGC 7145
QY 200 lGlnGluArgGlnLeuGlyArgProValPheProLys---AlaIleGlyGluLeuHis-- 218
Db 7146 CTACATCTGATGTCGCGACCTGCGCTGTCGCGAGCGCGCGCGCAACTCTACGT 7205
QY 219 -GlyHisThrGlyIleAlaArgLeuMetLeuPheTyAlaArgAlaAlaGly----- 236
Db 7206 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7259
QY 237 -----lleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyPh 254
Db 7260 GCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7307
QY 254 eAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyArgGlnValGlnIleThrPr 274
Db 7308 CACCGCGCGACCTG-----GTGCGCGCTGTGCGA 7334

RESULT 14
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4 (HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Alignment Scores:
Pred. No.: 0.0092 Length: 5163
Score: 123.50 Matches: 125
Percent Similarity: 34.60% Conservative: 85
Best Local Similarity: 20.59% Mismatches: 194
Query Match: 3.92% Indels: 203
DB: 35
```

US-09-758-269-6 (1-599) x US-08-700-651-1 (1-5163)

QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24
Db 2491 ACAATTGAGGATGTTGTTTCAGGA-----ATT 2517

QY 25 SerSerSerGlnSerSerAspLeuSerTyrcysSerSerLeu-----ProMetAlaSer 42
Db 2518 TCTGCAAGTGAGTCAATTA---TTATCTCAGAAATCAGCTCTAATCGACCCAGCAACAAAT 2574

QY 43 ArgValThrArgLys-----LeuAsnValSerSerAlaLeuHisThrPro 58
Db 2575 ATGGTGTGTTGGAGATTTGGTGGAATTTGTAACCCAGCAACAGGAGTGATGATTCAGGT 2634

QY 59 AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIle-----ValVal 74
Db 2635 TTTTAGGTGTCATCAGACAAACTCAATTCCTCCCTGAGATTGAAGATGGTGTGATTATT 2694

QY 75 LysProLysAlaLysGlnSerAsnThrLysGlnMetAsnLeu----- 88
Db 2695 CCTCCAGAAAGTACGACGACCAAAATGCTGATAAATTCAGTTATCTATTCCTCCAGCGTA 2754

QY 89 -----PheGlnArgAlaAlaAlaAlaLeuAsp 98
Db 2755 CCAGAAATCAATTCAGAAAGGATCAGAAAGATTGATTCTTAATTCGAATTGATGATGAT 2814

QY 99 AlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAsp 118
Db 2815 ATTGATGTCAGTACGATTTGTTGTCAGTATCAAGATCAAGACACCAATCCAGGTTC 2868

QY 119 ProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsn 138
Db 2869 -----ATTGCTGTGACTTGAACCCAAATTAATGAAGACACCAACACAACTGAC 2916

QY 139 LeuProValValGlyLysLeuProAspSerIleLysGlyValTyrrValArgAsnGlyAla 158
Db 2917 ---AGTGTACTGTAAACCAATCGATCCACACAGGTCTG----- 2955

QY 159 AsnProLeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHis 179
Db 2956 ---CCTTTCATCCACCA---ACTGGTCAAT-----TTGATTAAC 2988

QY 179 AlaValLysPheGluHisGlySerAlaSerTyr-----AlaCysArgPheThrGlnThr 196
Db 2989 CCACAAATTAATATACCATGGATCTCTTCATTTGCTGGTGCATACAAATATGCAATTCA 3048

QY 197 AsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGlu 216
Db 3049 AATCGTATTAACTGATGATAATGTTATGTTTACCAGTT-----GGTGAA 3093

QY 217 LeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrrAlaArgAlaAlaGly 236
Db 3094 ATAACAGGTTTACCAAGGATCCAGGCTCAGATATCCATTTAACTCACTACAGGTGAA 3153

QY 237 IleValAspProAlaHisGlyThrGlyValAlaAsn-----AlaGlyLeuValTyrrPhe 254
Db 3154 TTAGTTGATTCATCAACAGGAACCAATTAACATTTCTACTGCTGTTGTTGTT----- 3207

QY 255 AsnGlyArgLeu-----LeuAlaMetSerGluAspAspLeuProTyrrGlnValGlnIleThr 273
Db 3208 AGTGGAACCACTGGCTTACCACCTATTCAAGATGAA-----PheAsp 287

QY 274 ProAsnGlyAspLeuLysThrValGlyArgPheAsp-----TTTGATCCATCACTTCCCAATAGAT 3285

Db 3244 ---AATGGTAATTG----- 3285

QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
Db 3286 GGTAAAT-----AACCAATTAGTTAACCCAGAAACCAACAGCACT 3324

QY 308 PheAlaLeuSerTyrrAspValValSerLysProTyrrLeuLysTyrrPheArgPheSerPro 327
Db 327 ----- 327

Db 3325 GTCTCAGGATCAACTTCAGGTACTACAAAACCA-----AAACCA 3363

QY 328 -----AspGlyThrLysSerProAspValGluIle-----GlnLeuAsp 340

Db 3364 GGAATTCAGTCAATGGTGGAGGTGTTTACTGTGAAGAAGCTAAAGATCAAGCCGAT 3423

QY 341 Gln-----ProThrMet 344

Db 3424 AAGGGTAAGGATGATTAAATGTTCCACCACTAATTTCTATCAATAAAGATCCAGTAACA 3483

QY 345 MetHisAspPheAlaIleThrGluAsnPheValValProAspGlnValValPhe 364

Db 3484 AATACTCAGTACAGTAATACTACTGTGTAACATTAATTAACCCAGAAAGAAAGTT--- 3540

QY 365 LysLeuProGluMetIleArgGly-----GlySerProValVal 377

Db 3541 ---ATTCCAGGTTCACTTCCAGGCTCTCTCAACTATCCATCATTCATTAATCTCCACAA 3597

QY 378 TyrrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrrAlaGluAspSer 397

Db 3598 ACTGATGAG-----ATTACAGGAAGCCAGTGTGATCTGTACT 3636

QY 398 AsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGlu 417

Db 3637 GGTTCGCATAT----- 3648

QY 418 GluProGluThrAspGluValValIleGlySerCysMetThrProProAspSerIle 437

Db 3649 GATCCATCTCAGGTGAATTAATTCGATCTCGCAACTAAATTAATTCAGATCAGTT 3708

QY 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457

Db 3709 GCAGGTGAT-----GAAATCCTCACTGAAGTA---TTGAACATTACA 3747

QY 458 ThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGlu 477

Db 3748 ACAGATCAAGTAAACAGGTTTCCCAAT-----GATCTTGAA 3783

QY 478 AlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrrLeuAlaLeu 497

Db 3784 ACTGGTCTT----- 3792

QY 498 AlaGluProTrpProLysValSerGlyPheAlaLysVal-----AspLeuThrThr 514

Db 3793 -----CCAAGATCCAGTATCAGGACTCCCAACTTCCAAATGGTACTGTTGAT 3846

QY 515 GlyGluValLysLysHisLeuTyrrGlyAspAsnArgTyrrGlyGlyGluProLeuPheLeu 534

Db 3847 CCATCAATAAATAAACAACCAATTCAGGTTCA-----CATCCCGA-----TTTATT 3891

QY 535 ProGlyGluGlyGlyGlu 541

Db 3892 AATGGTACATCTCGAGAACAA 3912

RESULT 15
US-08-928-361B-4
Sequence 4, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSES: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verry, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1 (HV)
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

Alignment Scores:
Pred. No.: 0.0092 Length: 5163
Score: 123.50 Matches: 125
Percent Similarity: 34.60% Conservative: 85
Best Local Similarity: 20.59% Mismatches: 194
Query Match: 3.92% Indels: 203
Gaps: 35

US-09-758-269-6 (1-599) x US-08-928-361B-4 (1-5163)

QY	159	AsnProLeuHisGluProValThrGlyHisHisPhePheAspGlyMetValHis	178
DB	2956	---CCTTTCAATCCACCA---ACTGGTCAT	2988
QY	179	AlaValLysPheGluHisGlySerAlaSerTyr	196
DB	2989	CCACAAATAATAATACCATGGATTCTTCATTTGCTGTCATACAAATATGCAATTTCA	3048
QY	197	AsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaLeuGlyGlu	216
DB	3049	AATGGTATTAAAGCTGATAAATGTTTATGGTTTACAGTT	3093
QY	217	LeuHisGlyHisThrGlyLeuAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGly	236
DB	3094	ATAACAGGTTTACCAAGGATCCAGGCTCAGATATTCATTTAACTCAACTACAGGTGAA	3153
QY	237	IleValAspProAlaHisGlyThrGlyValAlaAsn	254
DB	3154	TTAGTTGATCCATCAACAGGAAGCCAAATTAACAATTTCTACTGCTGGTATTGTT	3207
QY	255	AsnGlyArgLeu---LeuAlaMetSerGluAspLeuProTyrGlnValGlnLeuThr	273
DB	3208	AGTGGAAACCTGGCTTACCACCTATTGAAGATGAA	3243
QY	274	ProAsnGlyAspLeuLysThrValGlyArgPheAsp	287
DB	3244	---AATGGTAAATTG	3285
QY	288	GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu	307
DB	3286	CGTAAAT	3324
QY	308	PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro	327
DB	3325	GTCTCAGGATCAACTTCAGGTACTACAAACCA	3363
QY	328	---AspGlyThrLysSerProAspValGluile	340
DB	3364	GGAATTCACAGTCAATGGTGGAGGTGTGTACCTGATGAAGAGCTAAAGATCAAGCCGAT	3423
QY	341	Gln	344
DB	3424	AAGGTAAGAGTGGATTAATTTGTTCCACCACTAATCTATCAATAAAGATCCAGTAACA	3483
QY	345	MetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPhe	364
DB	3484	AATACTCAGTACAGTAATACTACTGTGTACATTATTAAACCCAGAAACAGAAAGTT	3540
QY	365	LysLeuProGluMetIleArgGly	377
DB	3541	---ATTCCAGGTTCACTTCCAGGCTCTCTCAACTATCCATCATTCATCAATCAACAA	3597
QY	378	TyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSer	397
DB	3598	ACTGATGAG	3636
QY	398	AsnIleLysTrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGlu	417
DB	3637	GGTTTCCCATAT	3648
QY	418	GluProGluThrAspGluValValIleGlySerCysMetThrProAspSerIle	437
DB	3649	GATCCATCTACAGGTGAATATTCGATCTCGAATTAATTAATTAATTAATTAATTAATTAAT	3708
QY	438	PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys	457
DB	3709	GCAGGTGAT	3747
QY	458	ThrGlyLysSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGlu	477
DB	3748	ACAGATGAAGTAAACAGGTTTCCCAAT	3783
QY	478	AlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu	497

QY	5	ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProProLeu	24
DB	2491	ACAAATGAGGTAATGTTTTCAGGA	2517
QY	25	SerSerGlnSerSerAspLeuSerTyrCysSerSerLeu	42
DB	2518	TCGCAAGTGAAGTCAATTA---TTATCTCAGAAATCAGCTCTAATCGACCCAGCAACAAAT	2574
QY	43	ArgValThrArgLys	58
DB	2575	ATGGTGTGGAGAAATTTGGTGGATTGTGAACCCAGCAAGGAGTGTGATTCACAGGT	2634
QY	59	AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIle	74
DB	2635	TTTTTAGGTCCATCAGACAAACTCAATTCCTCCCTGAGATTGAAGATGGTGGTATTATT	2694
QY	75	LysProLysAlaLysGluSerAsnThrLysGlnMetAsnLeu	88
DB	2695	CCTCCAGAGTAGCAGCAGCAAACTGCTGATAAATTCAGTTATCTATTCCTCCAGCGTA	2754
QY	89	---PheGlnArgAlaAlaAlaLeuAsp	98
DB	2755	CCAGAAATCAATTCAGAAAGGATCAGAAAGTGTATTTCTGTAATGATGATGAT	2814
QY	99	AlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAsp	118
DB	2815	ATTGAGTCAAGTGAAGTATTATGCTCAAGTATCAAAAGAGACCAATCCAGGTCA	2868
QY	119	ProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsn	138
DB	2869	---ATTGCTGGTGAAGTGAAGCCCAATTAATGAAGACACCAACACAACTGAC	2916
QY	139	LeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAla	158
DB	2917	---AGTGTAACTGGTAAACCAATTCGATCCAAACACAGGTCTG	2955

Db	3784	ACTGGTCTT-----	3792
QY	498	AlaGluProTyrProlYsValSerGlyPheAlaLysVal-----	AspLeuThrThr 514
Db	3793	-----CCAAAGAGATCCAGATATCAGGACTCCACAACTTCCAAATGGTACCTTGGTTGAT	3846
QY	515	GlyGluValLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeu	534
Db	3847	CCATCAATAAATAAACCAATTCAGGTTCA-----CATTCCGGA-----	TTTATT 3891
QY	535	ProGlyGluGlyGlyGluGlu 541	
Db	3892	AATGTACATCTGGAGACAA 3912	

Search completed: November 17, 2003, 01:27:10
Job time : 8102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2003, 18:32:41 : Search time 431 seconds
(without alignments)
3751.655 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAAAGSRMLGNH.....VPYFGHTGIGADDLAKQVV 599

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=cgn2.1/UGPTO spool/US09758269/runat 14112003 192308 25751/app query.fasta_1.775
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09758269 @CGN_1.1 @runat 14112003 192308 25751 -NCPUS=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001C.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3150	100.0	1800	22	AAD09396	Arabidopsis thalia
2	2280.5	72.4	1818	22	AAD09401	Lycopodium escul
3	2168.5	68.8	1839	22	AAD09399	Vigna unguiculata
4	1991	63.2	1752	22	AAD09394	Arabidopsis thalia
5	1930	61.3	1815	22	AAD09400	Zea mays neoxanthi
6	1663.5	52.8	1734	22	AAD09398	Arabidopsis thalia
7	969	30.8	1788	21	AAC42989	Arabidopsis thalia
8	969	30.8	1788	21	AAC42989	Arabidopsis thalia
9	969	30.8	1788	24	AB213639	Arabidopsis thalia
10	960.5	30.5	1950	22	AAF77206	cDNA encoding sunf
11	939	29.8	1777	21	AA36083	Arabidopsis thalia
12	938	29.8	1617	22	AAD09397	Arabidopsis thalia
13	937	29.7	1617	22	AAD09402	Arabidopsis thalia
14	608	19.3	443	21	AAC56678	Eucalyptus grandis
15	592	18.8	492	22	AAC82706	Rice abscisic acid
16	556	17.7	491	21	AAC57157	Pinus radiata tran
17	522	16.6	393	25	ABX20484	Human GDP-mannose
18	511	16.2	372	21	AAC56695	Eucalyptus grandis
19	494.5	15.7	4403765	22	AA193683	Mycobacterium tube
20	494.5	15.7	4411529	22	AA193682	Mycobacterium tube
21	475.5	15.1	4403765	22	AA193683	Mycobacterium tube
22	475.5	15.1	4411529	22	AA193682	Mycobacterium tube
23	463	14.7	325	21	AAC56548	Eucalyptus grandis
24	415	13.2	386	21	AAC57162	Pinus radiata tran
25	409.5	13.0	340	21	AAC56520	Eucalyptus grandis
26	401	12.7	329	21	AAC57165	Pinus radiata tran
27	398	12.6	326	21	AAC57167	Pinus radiata tran
28	371.5	11.8	398	21	AAC57145	Human GDP-mannose
29	369	11.7	320	25	ABX20813	Pinus radiata tran
30	337.5	10.7	238	21	AAC57169	Pinus radiata tran
31	330.5	10.5	567	21	AAC56546	Eucalyptus grandis
32	322.5	10.2	1713	22	AA193635	Arabidopsis thalia
33	322	10.2	470	24	AB157937	Apoptosis inhibito
34	311	9.9	246	21	AAC57168	Pinus radiata tran
35	311	9.9	412	21	AAC57164	Pinus radiata tran
36	307.5	9.8	1791	22	AAD19634	Arabidopsis thalia
37	290	9.2	308	21	AAC57142	Pinus radiata tran
38	286.5	9.1	395	21	AAC57117	Pinus radiata tran
39	279	8.9	534	22	AAH44254	Physcomitrella pat
40	272.5	8.7	190	21	AAC57163	Pinus radiata tran
41	272.5	8.7	501	21	AAC43547	Zea mays DNA fragm
42	266.5	8.5	508	21	AAC56511	Eucalyptus grandis
43	260	8.3	3111	21	AAAS3888	Beta, Beta-caroten
44	256	8.1	1773	22	AAH42170	Nucleotide sequenc
45	252.5	8.0	447	21	AAC56480	Eucalyptus grandis

ALIGNMENTS

RESULT 1
AAD09396
ID AAD09396 standard; cDNA; 1800 BP.
XX AAD09396;
AC AAD09396;
XX
XX 10-SEP-2001 (first entry)
XX
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.
XX Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.
XX Arabidopsis thaliana.
OS
XX Key Location/Qualifiers
FH

Db 1321 TCTGACGAGATCTCAAGAGTGTCTCTCTCAATCGCCTGAATCTCAAAACCGGTGAA 1380
 Qy 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
 Db 1381 TCRACTCGCGTCGATCATCTCCACGAAAGATCAACAACTCAACCTCGAAGCAGGATG 1440
 Qy 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
 Db 1441 GTCACGAAACATGCTCGCGCTAAACCAATTCGCTTACCTTGGCTTAGCGAGCGG 1500
 Qy 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 1501 TGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCTACTCTGAGAGAGTTAAAGAACAT 1560
 Qy 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlu 540
 Db 1561 CTTTACGGCATACCGTTACGAGAGAGCCCTCTGTTTCTCCCGAGAGAGAGAGAG 1620
 Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
 Db 1621 GAACGAGAGATACATCTCTGTTCTGTTTCACGACGAGAGACATGGAATCGGATTA 1680
 Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 1681 CAGATAGTTAACGCGGTTAGCTTAGAGGTTGAAGCAACGGTTAACTTCCTCAAGGTT 1740
 Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspLeuAlaLysGlnValVal 599
 Db 1741 CCGTACGAGTTACCGGTACATTCATCGAGCCGATGATTTGGCAGACAGGTCGTG 1797
 RESULT 2
 AAD09401
 ID AAD09401 standard; cDNA; 1818 BP.
 AC AAD09401;
 XX
 XX 10-SEP-2001 (first entry)
 DT
 XX
 DE Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 cDNA.
 XX
 KW Tomato; neoxanthin cleavage enzyme; LeNCED1; abscisic acid; ABA;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant; herbicide; ss.
 XX
 OS Lycopersicon esculentum.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1818
 FT /*tag= a
 FT /product= "Lycopersicon esculentum LeNCED1 protein"
 XX
 EP1116794-A2.
 XX
 PD 18-JUL-2001.
 XX
 PF 11-JAN-2001; 2001EP-0300218.
 XX
 PR 13-JAN-2000; 2000JP-0010056.
 PR 11-JAN-2001; 2001JP-0003476.
 XX
 XX (RIKE) RIKEN KK.
 PA
 XX
 PI Iuchi S, Kobayashi M, Shinozaki K;
 XX
 DR WPI; 2001-400081/43.
 DR P-PSDB; AAE04789.
 XX
 XX A DNA encoding a protein with a neoxanthin cleavage activity for
 PT producing transgenic plants with improved or decreased stress tolerance
 PT
 XX
 PS Claim 3; Page 67-71; 101pp; English.
 XX

CC The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to a
 CC method for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant into which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes
 CC lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 protein
 CC related to the invention.
 XX
 SQ Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;

Alignment Scores:
 Pred. No.: 5,17e-213 Length: 1818
 Score: 2280.50 Matches: 435
 Percent Similarity: 82.04% Conservative: 63
 Best Local Similarity: 71.66% Mismatches: 90
 Query Match: 72.40% Indels: 19
 DB: 22 Gaps: 6

US-09-758-269-6 (1-599) x AAD09401 (1-1818)
 QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24
 Db 7 ACTACTCTTCACATGCCCAAAATACATGATTT-----AAGACTAAGTTG 51
 QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41
 Db 52 TCAATGCCATCATCAAGGAGTTGGTTTTCATCAAACTCTATTCTCTACTCAAAAT 111
 QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHis 61
 Db 112 CAACATAATAGCAAGTCTCAACATTAATTCCTCTCTCAAGCTCCACCTATCTCAT 171
 QY 62 PheProLysGlnSerSerAsn-----SerProAla-----IleValLysProLys 77
 Db 172 TTTCCTAAACATCTTCAAAATATCAACACCAAGAAATAATACAAATTTCAACCCAAA 231
 QY 78 AlaLysGluSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92
 Db 232 CAAGAAACACAACTCTCTCTCTCACTTCCAGTGGATTTAGTCAGAAAGCA 291
 QY 93 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112
 Db 292 GCAGCAATGGCTTTAGATGCTGTAGAAAGTGCTTTAACTAAACATGAACCTTGAACCCCT 351
 QY 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132
 Db 352 TTGCGAAACAGCCGACCCAGTCAGATTTCTGGGAATTTTGCTCCGGTACCGGAA 411
 QY 133 GlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal 152
 Db 412 AATCCAGTCTGTCAATCTCTCCGGTCCCGGAAATAATACCCAAATGTGTTCAGGCGTT 471
 QY 153 TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePheAsp 172
 Db 472 TAGGTTGAAACCGAGCTAACCTCTTTTGAACCAACCGCGGACACCAITTTCTGCAC 531
 QY 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg 192
 Db 532 GCGAGCGGTATGTTTACGCGCTTCAATTCAAAATGGTTCGGTGTAGTTACGTTGCCGT 591
 QY 193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212
 Db 592 TTCACGTGAACACAGAGAGGCTTGTTCAGAAAGAAAGTTTGGTTCGCCCTGTTTTCCTAAA 651
 QY 213 AlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArg 232

652 GCCATTGGTGAATTACATGCTCCTGGAATTCGAGGCTTATGCTGTTTACGCTCGT 711
233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuVal 252
712 GGGCTCTTCGGACTTGTGATCACAGTAAAGAACTGGTGTGTCAAACGCCGGTTAGTC 771
253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272
772 TATTTCAATCAACGATTAATCTGCTATGCTGAAAGATGATTTGCCTTACCATGTAAGGTA 831
273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292
832 ACACCCACCCGGCGATCTTAAACACAGAGGTCGATTCGATTTCCAGCGCCAGCTAAATCC 891
293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312
892 ACCATGATAGCTACCCCAAGCTCGACCCAGTTCCGGTGAGCTATTTCCTCTTAGCTAC 951
313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332
952 GATGTGATTCAGAGCCATACCTCAAGTACTTTCAGATTTTCAAAAAATGGGAAAAATCA 1011
333 ProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIleThrGlu 352
1012 AATGATGTTGAAATTCAGATTCAGAGCCCAACCAATGATGATGATTCGCAATTAAGT 1071
353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372
1072 AACTTCGTCGTCATCTGATCAACAAGTCGTTTTCAAGATGCTGAAATGATCCGTGA 1131
373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392
1132 GGTTCACCGTGGTTTACGACAGAACAAAGTTCCCGGATTTGGTATTCTGGATAGTAC 1191
393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeu 412
1192 GCGAAGATCGGTCCTGATTTGAATGGTTGAAGTACCTGATTTCTGTTTCCACCTC 1251
413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432
1252 TGGAAATGCTTGGGAAGACAGAGAACAGATGAATCGTTGTAATGCTTCATGTATGACA 1311
433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIle 452
1312 CCACAGACTCCATTTTCATGATGATGATGAGGCTTAAGAGTGTTCATCCGAAATC 1371
453 ArgLeuAsnLeuLysThrGlySerThrArgArgProIleLeuSerAsnGluAspGln 472
1372 CGTCTCAATTTGAAACAGGGAATCAACAGAAATCCATAATCGAAACCCCGGATGAA 1431
473 GluValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492
1432 CAAGTGAATTTAGAGCTGGATGGTGAACCGAACCAACTCGAAGGAACACAGAGTAT 1491
493 AlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512
1492 GCTTATTTGGTATCGCTGACCAACATGCGCAAAAGTTTCTGGTTTGCAGAAAGTAAACCTG 1551
513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532
1552 TTCACCGGTGAAGTTGAGAAATTCATTTATGGTGACAAACAAATGATGGTGGGAACTCTT 1611
533 PheLeuProGlyGlu-----GlyGlyGluGluAspGlyTyrIleLeuCysPheVal 550
1612 TTTTACCAAGAGACCCCAACAGCAAGCAAGAGACGATGTTATATTTTAGCTTCGTT 1671
551 HisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570
1672 CACATGAGAAGAAATGAAATCAGAACTGCAATTTGTTACGCAATGATGTTGAAATG 1731
571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590

1732 GAGGCAACTGTGAAGCTTCCATCAAGAGTTCTTATGATTTTCATGGAACATTCATAAAC 1791
591 AlaAspAspLeuAlaLysGln 597
1792 GCGATGATTTGGCAATCAG 1812

RESULT 3
AAD09399 standard; cDNA; 1839 BP.
XX AAD09399;
AC AAD09399;
XX 10-SEP-2001 (first entry)
XX Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 cDNA.
XX Cowpea; neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;
XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX plant growth protectant; CowPea Responsive to Dehydration; CPRD65; ss.
XX Vigna unguiculata.
OS
FH Key Location/Qualifiers
CDS 1..1839
FT /*tag= a
FT /product= "Vigna unguiculata CPRD65 protein"
XX
XX EPI116794-A2.
XX 18-JUL-2001.
XX 11-JAN-2001; 2001EP-0300218.
XX 13-JAN-2000; 2000JP-0010056.
XX 11-JAN-2001; 2001JP-0003476.
XX (RIKE) RIKEN KK.
XX Iuchi S, Kobayashi M, Shinozaki K;
XX WPI: 2001-400081/43.
XX P-FSDB; AAE04787.
XX
XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance
XX
XX Claim 3; Page 53-56; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
XX neoxanthin cleavage enzyme is introduced. The improvement of stress
XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
XX cleavage enzyme genes are useful for producing transgenic plants. An arid
XX land can be improved by growing transformant weed for several years and
XX then removing the weed by specifically lowering stress tolerance in the
XX weed by inducing an inducible promoter. The present cDNA sequence encodes
XX Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 (CowPea Responsive
XX to Dehydration) protein. CPRD65 gene is isolated from cowpea plant.
XX
XX Sequence 1839 BP; 447 A; 595 C; 442 G; 355 T; 0 other;

Alignment Scores:
Pred. No.: 4,84e-202 Length: 1839
Score: 2168.50 Matches: 423
Percent Similarity: 78.4% Conservative: 61
Best Local Similarity: 68.5% Mismatches: 98
Query Match: 68.8% Indels: 35

DB:	22	Gaps:	8
US-09-758-269-6	(1-599) x AAD09399	(1-1839)	
Qy	8	AlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeuSerSerSer	27
Db	7	TCATCAGCTTCAACACTTGGTTTAACGCCACACTCCCATCTCCCCCTTCAAGACCTA	66
Qy	28	GlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLys	47
Db	67	CTCTCCACATCT---TCTCCCAAACTTACTTCTCTTTA-----AGGAAACATCTCTCT	117
Qy	48	LeuAsn---ValSerSerAlaLeuHisThrProProAlaLeuHisPheProLysGln---	65
Db	118	TCCACACCATCACAATGTTTCCCTTCAACA-----CTCATTTCCCAACACGATAC	168
Qy	66	-----SerSerAsnSerProAlaIleValValLysProLysAlaLysGluSer	81
Db	169	CAACCAACATCCACATCCATCCACAGCCACCACCAACACCCACCCCAATCAAAAT	228
Qy	82	AsnThr-----	83
Db	229	ACCACCATCACCAACACACCGCCGAGGAAACCAACCTCTCTCTGACCAACCA	288
Qy	84	-----LysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAla	100
Db	289	CCATTACTCAAAATGGAATTTCTCCAGAAAGCGCTGCCACGSCCTTGGACCTGGTC	348
Qy	101	GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaSerProSer	120
Db	349	GAACCGCGCTGCTCTGCACGACGCAACACCCCTCCCCAAAACGGCGACCGAGG	408
Qy	121	ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro	140
Db	409	GTCCAAATCGCGGGAATTCGCGCGCGTCCCGAGCATGCCGAGCTCAAGACTCCCG	468
Qy	141	ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaSerPro	160
Db	469	GTGTCGGAATAATCCCAATGCATTGACGGGGTGTCGTGGCAACGGTGCATCCG	528
Qy	161	LeuHisGluProValThrGlyHisHisPheAspGlyAspGlyMetValHisAlaVal	180
Db	529	CTCTACAGAGCTGTGGCGGGCCACCTTCTTCGACGGGACGGCATGTCTCCACGGCTG	588
Qy	181	LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal	200
Db	589	AAGTTACGAAGCGCGCGCGAGCTACGCTCGCGCTTCAACGAGACGCGAGCTCTCTCG	648
Qy	201	GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis	220
Db	649	CAGAGAAATCTCTAGCGCGCGCGGTGTCCCGAAGGCCATCGGGAGCTCCACGGCCAC	708
Qy	221	ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspPro	240
Db	709	TCCGGCATCGCGGCTCTCTCTTACGCGCGGTCTCTTCGGGCTCGTGTATGGG	768
Qy	241	AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla	260
Db	769	TCCAGGGCATGGGCGTGGGAAACGCGGTCTCGTCTACTTCAACACCAACCTCTTGGCC	828
Qy	261	MetSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr	280
Db	829	ATGTCCGAGACGATTTACCTTACACGTGAGATCAACCTTAACCGCGACTTAACCAAC	888
Qy	281	ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal	300
Db	889	GTTGGCCCTTACGATTCACGGCGAGCTCAACTCAACATGATCGGCCACCCGAAATCG	948
Qy	301	AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu	320
Db	949	GACCCCGTCGCGGGAGCTTCAACGCGCTCAGCTACGAGCTCATCTCAGAGCCTTACTC	1008
Qy	321	LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp	340

QY 474 ValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAla 493
Db 1363 GTAATTTAGAGATTGGTATGTTTAAACCGGACCGTTAGAGAGAAAACCGGTTGCG 1422
QY 494 TyrLeuAlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAspLeuThr 513
Db 1423 TTTTGGCTATTGCTTATCCITGGCCAAAGTTTCGGTTTCGGTAAAGTCGATCTTTGC 1482
QY 514 ThrGlyGluValLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPhe 533
Db 1493 ACCGTGAGTGAAGAAATATATTACGGCGGTGAGAAATATGGCGGACACCGTTTTC 1542
QY 534 LeuProGlyGluGlyGly-----GluGluAspGluGlyTyrIleLeuCyaphe 549
Db 1543 TTGCCCGGCACTCCCGGTAAACCGCGAAGAAATGAAGATGACGGTTATATATTTTGTAC 1602
QY 550 ValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu 569
Db 1603 GTTCATGACGAGAAACAAAGACATCAGCTTCAGATTATTAACGCTGTTAATTTAAG 1662
QY 570 ValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIle 589
Db 1663 CTTGAAGCTACGATTAACTACCGTCTAGAGTACCGTATGGGTTCATGGCACATTGTG 1722
QY 590 GlyAlaAspAspLeuAlaLysGlnVal 598
Db 1723 GATTCGAATGAATCGTTGATCAATA 1749
RESULT 5
ID AAD09400 standard; cDNA; 1815 BP.
XX AAD09400;
AC AAD09400;
DT 10-SEP-2001 (first entry)
XX Zea mays neoxanthin cleavage enzyme, VP14 cDNA.
KW Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.
XX Zea mays.
XX Key Location/Qualifiers
XX CDS 1..1815
XX /tag= a
XX /product= "Zea mays VP14 protein"
PN BP1116794-R2.
XX 18-JUL-2001.
XX 11-JAN-2001; 2001EP-0300218.
XX 13-JAN-2000; 2000JP-0010056.
XX 11-JAN-2001; 2001JP-0003476.
XX (RIKE) RIKEN KK.
XX Tuchi S, Kobayashi M, Shinozaki K;
XX WPI; 2001-400081/43.
XX P-PSDS; AAE04788.
XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance
XX -
XX Claim 3; Page 60-64; 101pp; English.
XX The invention relates to neoxanthin cleavage enzymes and their

CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Zea mays neoxanthin cleavage enzyme, VP14 protein related to the
CC invention.
XX
SQ Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 other;
Alignment Scores:
Pred. No.: 1 05e-178 Length: 1815
Score: 1930.00 Matches: 375
Percent Similarity: 75.68% Conservative: 75
Best Local Similarity: 63.34% Mismatches: 122
Query Match: 61.27% Indels: 22
DB: 22 Gaps: 8
US-09-758-269-6 (1-599) x AAD09400 (1-1815)
QY 22 ProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMetAla 41
Db 49 CCGGCCCGGTCCAGGCCCGCGCTCC-----AATTCCTCAGGTTCTCG 93
QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrPro-----Pro 58
Db 94 CCGCGCGCGGTACAGTCCGTCGCGCGCGCGGTCTCCAGCGCGGTCCACAGGCC 153
QY 59 AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleValVal-----74
Db 154 GTCCGCGACCTGCTCGCGCTCCAGGAAGCCGCGCTTCGCCGCCAGCGCGCC 213
QY 75 ---LysProLysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAla 93
Db 214 CGCGCGCGAGGAAGCGGAGCGGCGGCAAGACAGCTCACTTGTTCAGCGCGCGCG 273
QY 94 AlaAlaAlaLeuAspAla---AlaGluGlyPheLeuValSer---HisGluLysLeuHis 111
Db 274 CGCGCGCGCTCGACCGTTCGAGGAAGAGGTTCGTGCCCAACGCTCCGAGCGGCCAC 333
QY 112 ProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsn 131
Db 334 GGGCTGCCAGCAGCGCGCGCGCGCGCGCGTGCAGATCGCGGCACTTCGCGCGCGCG 393
QY 132 GluGlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGly 151
Db 394 GAGAGCGCGCGCTGCACAGAGCTCCCGCTCCGCGCGCATCCCGCTTCATCGACGG 453
QY 152 ValTyrValArgGlnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePhe 171
Db 454 GTCTACGCGCGCAACCGCGCAACCCCTGCTTCGACCCCGTCCGCGGCGCACCACTTCT 513
QY 172 AspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAla---SerTyrAla 190
Db 514 GACGCGCAGCGCATGTGTGCACGCGTCCGCGATACGCAACGCGCGCGCGAGTCTCTAC 573
QY 191 CysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhe 210
Db 574 TGCGCGTTCAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 633
QY 211 ProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyr 230
Db 634 CCCAAGCCATTGGCGAGCTGCACGGGCACTCCGGGATCGCGCGCTCGCGCTGTTCTAC 693
QY 231 AlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGly 250

Db 694 GCAGCGCCGCGTGGCGCTGCTGGACCCCTCGCGCGGACCGCGGTGGCCAAACCCGCGC 753
Qy 251 LeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnVal 270
Db 754 CTGCTACTTCAACGCGCGCTGCTCGCATGCTCCGAGGACGACCTCCCTTACCACGCTC 813
Qy 271 GlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeu 290
Db 814 CGGTGGCGGACGACGCGGACCTCGAGACGCTCGCGCTGCTAGCACTTCAGCGGAGCTC 873
Qy 291 GluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeu 310
Db 874 GGCTGGCCCATGATCGGCACCCCAAGCTGGACCGCCGACCGGGAGCTCCACGCGTC 933
Qy 311 SerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThr 330
Db 934 AGCTACGACGTCATCAAGAGCGGCTACCTCACTACTTCTCACTCAGCGCCGACGCGACC 993
Qy 331 LysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIle 350
Db 994 AAGTCCGACGCTGGAGATCCGCTGGAGACGCCACCGATGATCCACGACTTCGCCATC 1053
Qy 351 ThrGluAsnPheValValProAspGlnGlnValPheLysLeuProGluMetIle 370
Db 1054 ACCGAGAACTCGTGTGTGTCGCGCACCCACCGGTGGTTCAGCTCCAGGAGATGCTG 1113
Qy 371 ArgGlyGlySerProValValTyrAspLysAsnLysValAlaAArgPheGlyIleLeuAsp 390
Db 1114 CGCGCGCGTGCCTCGTGTGCTGACAGAGAGACGTCGCGGTTCGCGGTGCTCCCC 1173
Qy 391 LysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPhe 410
Db 1174 AAGCACGCGCGGACGCGTGGAGATGGCGTGGGTGGACGTCGCGGACTGCTTCTGCTTC 1233
Qy 411 HisLeuTrpAsnAlaTrpGluProGluThrAspGluValValIleGlySerCys 430
Db 1234 CACCTGTGGAAACGCTGGAGAGACGCGGACGCGGCGAGGTGTGTGATCGCTCTGTC 1293
Qy 431 MetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSer 450
Db 1294 ATGACCCCGCGGACTCCATCTTCAACGAGTCCGACGAGCGCTGGAGAGCGTGTGACC 1353
Qy 451 GluIleArgLeuAsnLeuLysThrGlyLysThrArgProIleIleSerAsnGlu 470
Db 1354 GAGATCCGCTGGAGACGCGGACGCGGCGGTTCACGCGCGCGGCTTCCTGCGC---CCG 1410
Qy 471 AspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThr 490
Db 1411 TCGCAGCAGGAGAACTTGGAGTGGCATGTGTGAACCGCAACCTGCTGGGCGCGAGAGC 1470
Qy 491 LysPheAlaTyrLeuAlaLeuAlaGluProThrProLysValSerGlyPheAlaLysVal 510
Db 1471 CGGTACGCGTACCTCGCGGTGGGAGCGGTGGCCCAAGAGTGGGCTTCGCCAAGGAG 1530
Qy 511 AspLeuThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGlu 530
Db 1531 GACCTGTCCAGCGGCGGACTCACCAAGTTCGAGTACGCGGAGCGCGGCTTCGCGCGGAG 1590
Qy 531 ProLeuPheLeuProGlyGluGlyGlyGlu-----GluAspGluGlyTyr 545
Db 1591 CCCTGCTTCGTTCCATGACCGCGCGCGCGCCACCGCGCGCGGAGGAGCGGTAC 1650
Qy 546 IleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAla 565
Db 1651 GTGCTACCTTCGTCACGAGAGCGCGCGCGGACGCTCGGAGCTACTTGTGTCAATGCC 1710
Qy 566 ValSerLeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHis 585
Db 1711 GCCGACATCCGCTGGAGCGGAGTTCAGTGGCGTCCCGCGTCCCGCTTCGCTTCAC 1770
Qy 586 GlyThrPheIleGlyAlaAspLeuAlaLysGln 597
Db 1771 GGCACCTTCATCAGCGGCGGAGGCTCGAGGCCGAG 1806

RESULT 6

AAD09398 standard; cDNA; 1734 BP.

XX AAD09398;

XX 10-SEP-2001 (first entry)

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 cDNA.

XX Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 1..1734

FT /*tag= a

FT /product= "Arabidopsis thaliana AtNCED5 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX P-PSDB; AAE04786.

PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance

XX Claim 3; Page 46-49; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
XX neoxanthin cleavage enzyme is introduced. The improvement of stress
XX tolerance in plants is useful, for example in plant breeding. Neoxanthin
XX cleavage enzyme genes are useful for producing transgenic plants. An arid
XX land can be improved by growing transformant weed for several years and
XX then removing the weed by specifically lowering stress tolerance in the
XX weed by inducing an inducible promoter. The present cDNA sequence encodes
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.
XX The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA
XX library using a cDNA of the CPD65 (CowPea Responsive to Dehydration)
XX gene isolated from cowpea plant as a probe.

XX Sequence 1734 BP; 464 A; 390 C; 441 G; 439 T; 0 other;

XX Alignment Scores:

Score: 1.2e-152 Length: 1734
Pred. No.: 1663.50 Matches: 323
Percent Similarity: 72.56% Conservative: 100
Best Local Similarity: 55.40% Mismatches: 137
Query Match: 52.81% Indels: 23
DB: Gaps: 9

US-09-758-269-6 (1-599) x AAD09398 (1-1734)

PR 23-MAR-1999; 99US-0135788.
PR 25-MAR-1999; 99US-0136264.
PR 29-MAR-1999; 99US-0136785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132856.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147453.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

QY 531 roLeuPheLeuProGlyGluGlyGly-----GluGluAspGluGlyTyrIleL 547
 Db 1568 CGTTTTTCGTAGCTAGGATCCTGGTAATCCGAGCGGAGGAGGATGATGTTATGTGG 1627
 QY 547 euCysPheValHisAspGluGlyThrTrpLysSerGluLeuGlnIleValAsnAlaValS 567
 Db 1628 TGACGATGTTACGATGACTCGAGATCGAAGTTCTGGTGATGGACGCTAAAT 1687
 QY 567 er-----LeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheH 585
 Db 1688 CGCCGGAGCTTGAATCGTCGCGCCGTGAGGTTCGCGAAGGGTTCGTACGGATTCC 1747
 QY 585 iSGlyThrPheIleGlyAlaAspAspLeuAlaLys 596
 Db 1748 ATGGTTATTGTGTCAGGAAGTGACCTTAATAG 1782

RESULT 8

AA09395
 ID AAD09395 standard; cDNA; 1788 BP.

XX AC AAD09395;

DT 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 cDNA.

XX Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers
 FH 1..1788
 CDS /*tag= a

FT /product= "Arabidopsis thaliana AtNCED2 protein"

FN EP1116794-A2.

PD 18-JUL-2001.

PF 11-JAN-2001; 2001EP-0300218.

PR 13-JAN-2000; 2000JP-0010056.

PR 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX P-PSDB; AAE04783.

XX A DNA encoding a protein with a neoxanthin cleavage activity for
 PT producing transgenic plants with improved or decreased stress tolerance

XX Example 10; Page 25-29; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to
 CC methods for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant into which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein.

CC The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA
 CC library using a cDNA of the CPR65 (CowPea Responsive to Dehydration)
 CC gene isolated from cowpea plant as a probe.

XX SQ Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Alignment Scores:

Pred. No.: 1,21e-84 Length: 1788
 Score: 969.00 Matches: 214
 Percent Similarity: 55.31% Conservative: 114
 Best Local Similarity: 36.09% Mismatches: 221
 Query Match: 30.76% Indels: 44
 DB: 22 Gaps: 13

US-09-758-269-6 (1-599) x AAD09395 (1-1788)

QY 32 LeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLysLeuAsnValSer 51
 Db 56 CTCTTCTTCGCCGCGATCTTCTCTC-----CTACTC 88
 QY 52 SerAlaLeuHisThrProProAlaLeuHisPheProLysGlnSer-----SerAsn 68
 Db 89 TCTCCGTATCAACTCCGCGTCTCGAAGACGTTCTCCAATCACAACCCAGCAGACA 148
 QY 59 SerProAlaIleValLysProLys-AlaLysGluSerAsnThrLysGlnMetAsnLe 88
 Db 149 ACAATGATCGGTGTAACAACCCAAACACTCCCAACCGCAACCAATCACACCTTAGTCT 208
 QY 88 uPheGlnArgAlaAlaAlaAlaLeuAspAla-AlaGluGlyPheLeuValSerHisG 108
 Db 209 CATCACACCCGAAACTCCGACCAGAAATGACTCTCGCAACAGACTCTCTCCACCCGCTG 268
 QY 108 luLysLeu-----HisProLeuProLysThrAlaAspProSerValG 122
 Db 269 AAGATGTAATCAACACACGTTTCATCGATCCACTTCACGTCCTTCGTTGATCCAAACATG 328
 QY 122 InileAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValV 142
 Db 329 TCCITCTCTGATACTTCGCTCTCTCTCGACGAGCTTCTCCACAGACTGTGAAATCA 388
 QY 142 al---GlyLysLeuProAspSerLeLysGlyValTyrValArgAsnGlyAlaAsnProL 161
 Db 389 TCCACGGCACTCTTCCACTGTCACTTAACGGCGCTTACATCCGTAACGGTCCAAATCCAC 448
 QY 161 euHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValL 181
 Db 449 AGTTTCTCCCTCGTGCTCTTACCATCTCTTCGACGGCGAGGTATGCTTCACGCCATAA 508
 QY 181 ysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValG 201
 Db 509 AAATCCACAACGGTAAGCCACTCTCTGTAGCAGATAGCTCAAGACTTATAAATACACG 568
 QY 201 lngluArgLysGluGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyH 221
 Db 569 TCGAACAACAACCGGAGCTCGGTATGCTTAACGTTTTCGGGATTCACGGTGTA 628
 QY 221 hr---GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAsp 240
 Db 629 CGGCGTCAGTAGCTCGTGAGCTTTAACGGCAGCTAGGGTTTAAACCGCAGATATAATC 688
 QY 240 roAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeu 260
 Db 689 CGGTAAACGGCATTCGTTTAGCTAATACAGTCTAGCTTTCTTCTAGTAACCGTCTCTTTG 748
 QY 260 laMetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysT 280
 Db 749 CTTTAGGTGAATCTGATTACCTACCGCGTCCGATTAAACCGAATCAGAGATATTGAAA 808
 QY 280 hrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysV 300
 Db 809 CGATCGGAGCGGATTCGAGGAAATAGCGATGAGTATGACAGCTCATCTCTAATAA 868
 QY 300 alAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrL 320

```
Db 869 CCGATCCAAACACGAGAACTTTCCGTTCCGTTACGTCGCGTTT 925
Qy 320 euLysTy-PheArgPheSerProAspGlyThrLysSerProAspValGluLeu---GlnL 339
Db 926 TAAATATTTCCGTTTGAATCCCGCGGAAACAAAGAGACGTTCCGATATTCGGA 985
Qy 339 euAspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProA 359
Db 986 TGACGCTCCGTCGTTTCCCATGACTCCGATCACGAAAGCTCACGCGATTTCGCAG 1045
Qy 359 spGlnGluValValPheLys-----LeuProGluMetIleArgGlyGlySerProV 376
Db 1046 AGATTCCAGTTCCGATGAGGATGAACATGTTGGATTGTTCTCGAAGTGGTTCTCCGG 1105
Qy 376 alValTy-AspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyValaGluAsps 396
Db 1106 TTGGTACTGATAACCGAAACCTCAAGCTTGGAGTGCATTCCTAAGTACCGCGAGATG 1165
Qy 396 exSerAsnIleLysTrpIleAspAlaProAspCysPheHisLeuTrpAsnIle 416
Db 1166 ATCTCGAGATGAATGGTTCGAAGTTCCTGGATTCAATATATTCACGCTATTATGCTT 1225
Qy 416 rpGluGluProGluThrAspGluValValValIleGlySerCysMetThrProProAsps 436
Db 1226 GCGATGAAGATCATGAACACGCTGTTTTCATTCACCGAATATATTCGATTGAAC 1285
Qy 436 erIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnL 456
Db 1286 ATACTTTAGAGAGATGATGAT---CTGGTTCATGCTTTGGTGAAGATGAGATCGATC 1342
Qy 456 euLysThrGlyGluSerThrArgAspProIleSerAsnGluAspGlnValAsnL 476
Db 1343 TCGTACCGGATTTGAGACGTCATCGATCTCAGCGAGG-----AATC 1387
Qy 476 euGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuA 496
Db 1388 TCGATTTCGCTGTGATTAATCCGGGCTTCTCGGAGATGTAGCAGGTACGTTACGCGG 1447
Qy 496 laLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyG 516
Db 1448 CGATTGGAGATCCGATGCGGAAGATCTCCGGTGTGTTGAGCTTGTATGTTAAAGAG 1507
Qy 516 lu-----ValLysLysHisLeuTyArgAspAsnArgTyArgLysGlyGluP 531
Db 1508 ATCGGATGATTGACGTTGGCCCGCTAGATGATGACGTTTCAGTTGTTACGCGGAGAAC 1567
Qy 531 roLeuPheLeuProGlyGlyGly-----GluGluAspGlyTyTrileL 547
Db 1568 CATTTCCTAGCTAGGATCTGTAATCCGAGCGGAGGAGGATGATGTTATGTTGG 1627
Qy 547 euCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValS 567
Db 1628 TCACGTATGTTCCAGCATGAAGTGACTGGAGATCGAAGTTTCTGGTATGACGCTAAT 1687
Qy 567 ex-----LeuGluValGluAlaThrValLysLeuProSerArgValProTyGlyPheH 585
Db 1688 CCGCGAGCTTGAATTCGTCGCGCGGTGAGGTTCGCGAAGGTTCCGACGAGTTCC 1747
Qy 585 isGlyThrPheIleGlyAlaAspAspLeuAlaLys 596
Db 1748 ATGGGTATTCTCAGGAAAGTACCTTAATAG 1782
```

RESULT 9

AB213639

ID AB213639 standard; DNA; 1788 BP.

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

XX AC AB213639;

KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN WO200216655-A2.
XX 28-FEB-2002.
PD 24-AUG-2001; 2001WO-US26685.
XX 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-284647P.
XX 22-JUN-2001; 2001US-300111P.
XX (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
PI Harper JF, Kreps J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
DR WPI; 2002-304127/34.
XX Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT -

Claim 144; SEQ ID NO 1444; 577bp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:

(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and

(b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used in methods of the invention.

Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

XX Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Alignment Scores:

Pred. No.: 1-21e-84 Length: 1788

Score: 969.00 Matches: 214

Percent Similarity: 55.31% Conservative: 114

Best Local Similarity: 36.09% Mismatches: 221

Query Match: 30.76% Indels: 44

DB: 24 Gaps: 13

US-09-758-269-6 (1-599) x AB213639 (1-1788)

Qy 32 LeuSerTyrcysSerSerLeuProMetAlaSerArgValThrArgLysLeuAsnValSer 51

Db 56 CTCTTCTCCGCGCGATCTCTCTCTC-----CTACTC 89

Qy 52 SerAlaLeuHisThrProAlaLeuHisPheProLysGlnSer-----SerAsn 68

Db 89 TCCTCGTATCACTCCGCGCTCGAAGACGTTCTCCATCAACACCCAGCGACA 148

Qy 69 SerProAlaIleValLysProLys-AlaLysGluSerAsnThrLysGlnMetAsnLe 88

Db 149 ACAATGATCGTCGTAAACAAACCAACCACTCCACAAACCAACCACTACACCTTAGTCT 208

Qy 88 uPheGlnArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 108

Db 209 CATCACCCGGAACCTCCGACCAAGAAATGACTCTCGCAACAGCTCTCTTCACCCAGCTG 268

Qy 108 lylLysLeu-----HisProLeuProLysThrAlaAspProSerValG 122

Db 269 AAGATGTAATCAACACGTTTCATCGATCCACCTTCACGCTCCGTTGATCCAAACATG 328

PS Claim 1; Page 94-97; 135pp; English.

XX This invention relates to defence-related signalling genes isolated from the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich protein (GRP). The signalling gene is useful for increasing the resistance of a plant to a pathogen such as fungus, virus, bacterium, nematode or insect (e.g. European corn borer), preferably *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a construct containing the gene into the genome of the plant. The gene is useful for regulating gene expression in a plant, in response to a stimulus such as infection with a pathogen, damage from a pathogen, hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid, oxalic acid or expression of a gene encoding oxalic acid oxidase. The genes are also useful for stem-preferred regulation of gene expression in a plant. The genes are useful in agriculture, particularly in the breeding of crop plants with improved agronomic traits, for modifying abscisic acid (ABA) metabolism and for modifying amino acid transport and content in plants. The present sequence represents cDNA encoding the sunflower neoxanthin cleavage enzyme (NCE).

XX Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Alignment Scores:

Pred. No.:	9.37e-84	Length:	1950
Score:	960.50	Matches:	209
Percent Similarity:	53.9%	Conservative:	109
Best Local Similarity:	35.48%	Mismatches:	214
Query Match:	57	Indels:	57
DB:	22	Gaps:	13

US-09-758-269-6 (1-599) x AAF77206 (1-1950)

QY 51 SerSerAlaLeuHisThrProAlaLeuHisPheProLysGlnSerSerAsnSerPro 70
 DB 49 TCNACTGTTCCACCACACCA-----TCATCGTCTCACCG 87

QY 71 AlaileValVal-----LysProLysAlaLysGlnSer 81
 DB 88 CCATCTCGGTAAAGTCTTTTCAGTCAGACTGAAGAAACACAGAACCGTCCACG 147

QY 82 -----AsnThrLysGln-----MetAsnLeuPheGln 90
 DB 148 ACGCCACACCAACCAAGACCTAGTGACCGAAACCAACCAACCTGAATATTAGAA 207

QY 91 ArgAlaAlaAlaAlaLeuAspAla-----AlaGluGly 102
 DB 208 CGTGAAGTTCAGTGTGGTGGATCAGTCTTTACCGCCACCATCTTTAATGCTTTGAT 267

QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
 DB 268 AGCATATAATAACTTTTATTGATCCACCGGAGGGTCTCAGTTGATCCAAACACGTT 327

QY 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142
 DB 328 TTCTCTGATAACTTTTCACCGGTGGACGAACTCCCTCCGACTGACTGTGAAGTCATCGAG 387

QY 143 GlyLysLeuProAspSerLeuLysGlyValThrValArgAsnGlyAlaAsnProLeuHis 162
 DB 388 GGCACACTGCCAAGTTCCTTTCAGTGTGGTGTACTTCCGTAATGGTCCGAACCCGCAATTC 447

QY 163 GluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLysPhe 182
 DB 448 CTTCCCGAGGACCTTACCACTCTTCGATGGGATGGCATGCTCATGCTATTGATC 507

QY 183 GluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGlu 202
 DB 508 TCCAATGGAAAGCTTCGTTATGTAGCCGATACATCAAAACATACAAATATTCAATAGAG 567

QY 203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisHisThr--- 221
 DB 568 AAGAAGCGGGATGCCCATATTTCCAAACGCTTTTACAGGTTTAATGTGTGACTGCC---627

QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAla 241
 DB 628 TCTGCAGCTCGCATGCACTCACTCGCGCGCATTTTGGCTGGACAATTTGACCCACCA 687

QY 242 HisGlyThrGlyValAlaAlaAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMet 261
 DB 688 AARGGTATTGGCTAGCCAAATACAGTCTGGCTTTTGGCAACAGACTTTTGTCTGTG 747

QY 262 SerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281
 DB 748 GGAGAGTCGGATCTCCCATATGCGTCGCAAACTAGCGCCGCGGTCATAGTACCGGTC 807

QY 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301
 DB 808 GGAGTCGAGACTTCGACGGCAACATTTTCATGAGCATGACCTCACCACCAAAATCGAT 867

QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys 321
 DB 868 CCAGTAACGAAGAAGCTTTTGGCTTTTGGTACGGTCCAGTC---CCCCCTTTCTTAACC 924

QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluIle---GlnLeuAsp 340
 DB 925 TTTTTCGGTTTCAACGAAACCGAGAAACAAACGCGCATGTCGCCGATCTTCTCAATGACA 984

QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
 DB 985 AGCCCGTCGTTCTCCACGACTTCGCCATCACCAAACTACCGGATTTTCCGAGATC 1044

QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
 DB 1045 CAAATCGGATGAGCCCAATGCGAGATGCTGGGTGGGGATCCCGGTTAGCGGCGGCT 1104

QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
 DB 1105 GGAAGGTGCTCGGTCGGTGTGATCCCTCGTACGCGAAACGAGTCCGAGATGAAG 1164

QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
 DB 1165 TGGTTTCAAGTTCGGGTTTTTAATGTATCATTCATCATCATCATCATCATCATCATCAT 1224

QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
 DB 1225 GGAGATACGGT 1284

QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 DB 1285 ATGGAT---TTGATTCATGATCGTTCGTTAGAAAAGTGAAGATTAACTCTGAAAACAGGAATG 1341

QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
 DB 1342 GTATCCCGGACCCGCTTTCACCCCGG-----AATCTTGCTTTGGCGTT 1386

QY 481 ValAsnArgAsnMetLeuGlyArgGlyThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
 DB 1387 TTAATCCGCTTTGTTGCGCTTAAACACAGGTATATGTTGTGGAGTTGGTGTATCCG 1446

QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGlu----- 516
 DB 1447 ATGCCAAAGATCTCTGGTGTGGTCAAGTAGATGTGCTACTCCGAGCGAGCCGCTCGC 1506

QY 517 -----ValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPhe 533
 DB 1507 GAATGCATAGTTGTCAGCCGATGTTTCGGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1566

QY 534 LeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIleLeuCysPhe 549
 DB 1567 GTAGCTAGGAGCGCAGCAACCCCGATCGCGATGAGGATGAGGATGAGGATGAGGATGAGG 1626

QY 550 ValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSer----- 567
 DB 1627 GTGATACAGCAACACCGGTGAGTCAAGATTGCTGGTGTGAGCGCTAAGTACCCGACG 1686

QY 568 LeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThr 587

```
Db      1687 CTTGAGATCGTGCATCCGTGAAGCTGCCCCACCGGTACCATATGGTTTTCATGGGTC 1746
Qy      588 PheileglyAlaaspAspLeualllys 596
Db      1747 TTGTGAGAGAAAGTGACCTTAACAG 1773

RESULT 11
AAC36083
ID AAC36083 standard; DNA; 1777 BP.
AC AAC36083;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12487.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135623.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
```



```

US-09-758-269-6 (1-599) x AAD09402 (1-1617)
QY 64 LysGlnSerSerAsnSerProAlaIleValValysProLys---AlaLysGluSerAsn 82
DB 10 AAACCTCAGTATGCGCAGCATCATCTCAGTCCATCTAGACCCCTCCCAAGGTTCTCC 69
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaAlaAlaAlaGluGly 102
DB 70 TCGAAGCTTCTCGATCTCTCGAGAGACTTGTCTCAAGCTCATG-----114
QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAlaAspProSerValGln 122
DB 115 -----CACGATGCTTCTCTCCCTCTCCAC-----TAC 141
QY 123 IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgAsnLeuProVal 141
DB 142 CTCTCAGGCAACTTCGCTCCCATCCCGTGATGAACTCCCTCCGTCAGGATCTCCCGTC 201
QY 142 ValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAlaAsnProLeu 161
DB 202 CATGGATTCTTCCCGATGCTTCAATGGTGAATTTGTGAGGTTGGTCCAAACCCCAAG 261
QY 162 HisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLys 181
DB 262 TTTGATGCTGCTCGCTGATATCATCTGTTTGTGAGATGGGATGATTCATGGGTACGC 321
QY 182 PheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGln 201
DB 322 ATCAAGATGGGAAGTACTTATGTTCTCGATATGTTAAGACATCACGCTTTAAGCAG 381
QY 202 GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221
DB 382 GAAGAGTTCTTCGGAGCTGCGCAATTCATGAAG---ATTGGTGACCTTAAGGGGTTTTC 438
QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAla 241
DB 439 GGATTTGCTAATGGTCAATATCCACAGCTGAGAGAGAGTCAAAATATTGGCAACACT 498
QY 242 HisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMet 261
DB 499 TATGGAATGGAAGTGGCAATACAGCACTCGTATATCACCATGAGGAAACTTCTAGCATTA 558
QY 262 SerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281
DB 559 CAGAGGCGAGATAGCCGCTAGTCAAAAGTTTGGAAAGATGGAGACCTGCAAACTCT 618
QY 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301
DB 619 GGTATAATAGATTATGACAAGAGATTGACCCACTCTCTCACTGCTCACCCCAAAAGTTGAC 678
QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys 321
DB 679 CCGGTACGGGTGAATGTTTACATTCGGCTATTTCG---CATACGCCACCTTATCTCACA 735
QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341
DB 736 TACAGAGTTATCTCGAAGATGGCATTATGATGATCCCGAGTCCCAATTACTATATCAGAG 795
QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361
DB 796 CCTATCATGATGATGATGATTTGCTATTACTGAGACTTATGCAATCTTCATGATGATCTTCT 855
QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValVal---TyrAspLys 380
DB 856 ATGCATCTCAGCCCAAGGAATGGTGAAGAGAAGAAATGATATCTATCTATTTGATCCC 915
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
DB 916 ACAAAGAGGCTCGTTTGGTGTCTCTCCGGCTATGCCAAGGATGAACATTATGATTAGA 975
QY 401 TrpIleAspAlaProAspCysPheHisLeuTyrAsnAlaTyrGluGluProGlu 420

```

```

DB 976 TGGTTTGAGCTTCCCACTGCTTTATTTTCCACAAAGCCCAATGCTTGGGAA-----GAA 1029
QY 421 ThrAspGluValValIleGlySerCysMetThrProProAsp-----435
DB 1030 GAGGATGAAGTCGCTCCATCATCTGCTTGGAGAAATCCAGATCTTGACATGGTCACT 1089
QY 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455
DB 1090 GGGAAAGTGAAGAAAGAACTCGAAATTTTGGCAACGAACCTGTACGAAATGAGATTCAAC 1149
QY 456 LeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsn 475
DB 1150 ATGAAACGGGCTCAGCTTCTTCAAAAAAATACTCCGCATCTCGC-----1194
QY 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeu 495
DB 1195 GTTGAATTTCCCAAGATCAATGAGTGTACACCGGAAGAAACAGAGATACGTATATGGA 1254
QY 496 AlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAsp-----511
DB 1255 ACAATTTCTGGACAGTATCGCAAGGTTTACCGGAATCATCAAGTTTGTATCGCATGCGAGAA 1314
QY 512 -----LeuThrThrGlyGluValLysLysHisLeuTyr-----522
DB 1315 GCTGAGACAGGAAAGAAAGTCTCGAAGTAGGAGGTAAATATCAAGGAATATATGACCTG 1374
QY 523 GlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGluGlyGlyGluGluAsp 542
DB 1375 GGAGAAAGCAGATATGTTTCAGAGGCTATCTATTTCCGCTGAGACAGCAGAAAGAGAC 1434
QY 543 GluGlyTyrIleLeuCysPheValHisAspGluLysThrTyrLysSerGluLeuGlnIle 562
DB 1435 CAGCGTTTACTTGATATTTCTTTTCATGATGAAACACAGGGAATATCATGCGTCACTGTG 1494
QY 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580
DB 1495 ATAGACGCAAAACAAATCTCGCTCAACCGGTGGCAGTGTGGAGTCCCGCAGGAGGTC 1554
QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
DB 1555 CCATATGCTTCCATGCGCTTGTGTTTGTACAGAGAAACAACTCCAGGAACAACTCTT 1611

RESULT 14
AAC56678
ID AAC56678 standard; DNA; 443 BP.
AC AAC56678;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #549.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Sherk MA, Glenn M;

```

XX DR WPI; 2000-579369/54.
 XX XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 XX
 XX
 PS Claim 1; Page 490; 747pp; English.
 XX
 XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-Box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/VADS, homeodomain zipper, LIM domain, AP2
 CC and ERBPs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 443 BP; 71 A; 177 C; 117 G; 77 T; 1 other;
 Alignment Scores:
 Pred. No.: 3.38e-50 Length: 443
 Score: 608.00 Matches: 115
 Percent Similarity: 87.76% Conservative: 14
 Best Local Similarity: 78.23% Mismatches: 18
 Query Match: 19.30% Indels: 0
 DB: 21 Gaps: 0
 US-09-758-269-6 (1-599) x AAC56678 (1-443)
 QY 208 ProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyValAlaArgLeuMet 227
 Db 3 CCGCTCTCCNCAAGCCATCGCGAGCTCCACGCCCACTCCGGCATCGCGGCTCATG 62
 QY 228 LeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAla 247
 Db 63 CTCCTCTACGCCCGGAGCTCTTCGGCTCGTCCGACCCGGAATGGCATGGCGTCCGG 122
 QY 248 AsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267
 Db 123 AACGCCGGCTGCTGACTTCGACGCCACCTCTTCGGATGTCGAGGACGACCTCCCC 182
 QY 268 TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp 287
 Db 183 TACCAGTCGGGTACCGGCTCCGGCGACCTCGAGACCGTCCGGCGCTACGACTTCGCC 242
 QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
 Db 243 GGCCAGCTCGATCTCCGATGATGCCCCACCCGAGATCGACCCGGCTTCGGCGAGATG 302
 QY 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327
 Db 303 TTCGCCCTCAGCTACGACGTCGTCGGAAGCGGTACCTCAAGTACTTCGATCTCCCAAG 362
 QY 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMethIleAsp 347
 Db 363 GACGGGGAAGTCCCGGACGTCCGATGCCCTTCGATGCCCTTCGGTGTGACCCGACCATG 422
 QY 348 PheAlaIleThrGluAsnPhe 354
 Db 423 TTCGCCATCACCAGCGCTTT 443
 RESULT 15
 AAC82706/c
 ID AAC82706 standard; DNA; 492 BP.
 XX
 AC AAC82706;
 XX

DT 15-MAR-2001 (first entry)
 XX Rice abscisic acid synthesis associated DNA SEQ ID NO 5.
 DE
 XX Rice; abscisic acid synthesis; VP14-like gene; drought resistance; ds.
 XX
 OS Oryza sativa.
 PN WO2000071727-A1.
 XX
 XX 30-NOV-2000.
 PD
 XX 25-MAY-1999; 99WO-JP02734.
 PF
 XX 25-MAY-1999; 99WO-JP02734.
 PR
 XX (NORU) JAPAN MIN AGRIC FORESTRY & FISHERIES
 PA (BIOOC-) BIO-ORIENTED TECHNOLOGY RES ADVANCEMENT.
 EA Hirochika H, Sakamoto K;
 XX WPI; 2001-032042/04.
 XX
 PT Oligonucleotide encoding gene for regulating abscisic acid synthesis in
 PT plants, useful for constructing e.g. genetically-modified rice with
 PT drought resistance and ear-germination resistance
 XX
 PS Disclosure; Page 50; 55pp; Japanese.
 XX
 CC This invention describes a novel polynucleotide sequence (I) which
 CC encodes a protein capable of regulating the synthesis of abscisic acid.
 CC The invention also describes (1) an oligonucleotide encoding a protein
 CC of a gene obtained by controlling the expression of a VP14-like gene;
 CC (2) a vector containing the oligonucleotide ligated operably to the
 CC regulation sequence; (3) a plant transformed with the vector; and
 CC (4) a method for regulating abscisic acid synthesis in a plant including
 CC the transfer of the above oligonucleotide to it. The gene is useful for
 CC constructing drought resistant rice.
 XX
 SQ Sequence 492 BP; 90 A; 126 C; 186 G; 88 T; 2 other;
 Alignment Scores:
 Pred. No.: 1.45e-48 Length: 492
 Score: 592.00 Matches: 115
 Percent Similarity: 90.28% Conservative: 15
 Best Local Similarity: 79.86% Mismatches: 13
 Query Match: 18.79% Indels: 2
 DB: 22 Gaps: 0
 US-09-758-269-6 (1-599) x AAC82706 (1-492)
 QY 210 PheProLysAlaIleGlyGluLeuHisGlyHisThrGlyValAlaArgLeuMetLeuPh 229
 Db 434 TTTCCTAAGCGGATAGTGAGCTCCATGGCCACTCCGGGATCGCGGCTGCTGTGT 375
 QY 229 eTyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAla 249
 Db 374 CTACGGCGGGCGGCTC-GGCTCTCCGACCGGTCCACACGCGCACCGGCTGCCAACGC 316
 QY 249 aGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrG 269
 Db 315 CGGCCCTCATCTTCAACGGCAGGCTCTTCGGCATGTCCGAGGACGACCTCCCTTACCA 256
 QY 269 nValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGly 289
 Db 255 GGTGGCGGTCAACCGCGGCGGACCTCGAGACCGTCCGCGGTACGACTTCGACGGGCA 196
 QY 289 nLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPhe 309
 Db 195 GCTCGGCTGGCCATCATCGGCACCCCAAGTCGACCGCGGCGGACCGGAGAGCTCCACGC 136
 QY 309 aLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAsp 329

Mon Nov 17 10:00:59 2003

Db 135 GCTCAGCTACGACGTGATCAAGAAGCCGTACCTCAAGTACTTCTTCTCGCCCGACGG 76
Qy 329 yThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAl 349
Db 75 CACCAAGTCGGCGGACGTCGAGATCCCGCTCGACCCACCATGATCCACGACTTCGG 16
Qy 349 alleThrGlu 352
Db 15 AATTACTGAG 6

Search completed: November 16, 2003, 20:34:06
Job time : 477 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 16, 2003, 20:09:17 ; Search time 6134 Seconds
(without alignments)
3994.924 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASPTATAAAGSRWLGNNHT.....VPYGFHGTFIGADDLAKQVV 599

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09758269/runat_14112003_192309_25759/app_query.fasta_1.775
-DB=GenEmbl -CPMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09758269 @CGN_1_1_3963 @runat_14112003_192309_25759 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.ov.*
5: gb.cv.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rnd.*
36: em.htg_man.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3150	100.0	1800	6	AX148306 Sequence
2	3150	100.0	1800	6	BD017431 Transgeni
3	3150	100.0	3869	8	AB026549 Arabidops
4	3150	100.0	52232	8	AB028617 Arabidops
5	3143	99.8	2331	8	AY056255 Arabidops
6	2291	72.7	2164	8	AJ276244 Solanum t
7	2282.5	72.5	2171	8	Z97215 Lycopersico
8	2282.5	72.5	19018	8	AJ439079 Lycopersi
9	2280.5	72.4	1818	6	AX148316 Sequence
10	2280.5	72.4	1818	6	BD017436 Transgeni
11	2172	69.0	2398	8	AF190462 Phaseolus
12	2170.5	68.9	95769	6	AX652128 Sequence
13	2170.5	68.9	95769	6	AC013430 Genomic s
14	2168.5	68.8	1839	6	AX148312 Sequence
15	2168.5	68.8	2432	6	BD017434 Transgeni
16	2168.5	68.8	2432	6	AB030293 Vigna ung
17	2155.5	68.4	2357	8	AB080192 Pisum sat
18	2075.5	65.9	2034	8	AB080193 Pisum sat
19	2066.5	65.6	2310	8	AF224671 Persea am
20	2044.5	64.9	72058	8	AC074176 Arabidops
21	1991	63.2	1752	6	AX148302 Sequence
22	1991	63.2	1752	6	BD017429 Transgeni
23	1991	63.2	110102	8	ATF28J12 Arabidops
24	1991	63.2	194143	8	ATCHRIV48 Arabidops
25	1988	63.1	175158	2	AC120531 Oryza sat
26	1988	63.1	181159	2	CNS07YP7 Oryza sat
27	1987	63.1	2153	8	AF224672 Persea am
28	1962	62.3	133983	8	CNS08CB9 Arabidops
29	1959.5	62.2	131689	8	CNS08CB9 Arabidops
30	1930	61.3	1815	6	AX148314 Sequence
31	1930	61.3	1815	6	BD017435 Transgeni
32	1930	61.3	2498	8	ZM095953 Zea mays vi
33	1834.5	58.2	149097	2	AF005632 Oryza sat
34	1663.5	52.8	1734	6	AX148310 Sequence
35	1663.5	52.8	1734	6	BD017433 Transgeni
36	1663.5	52.8	78921	8	AB028621 Arabidops
37	1214	38.5	82697	8	ATT9A21 Arabidops
38	1014.5	32.2	120055	8	AC123596 Medicago
39	1003	31.8	80686	2	AC144759 Medicago
40	1003	31.8	120055	8	AC123596 Medicago
41	986.5	31.3	2016	8	AB080194 Pisum sat
42	971.5	30.8	80686	2	AC144759 Medicago
43	969	30.8	1788	6	AX148304 Sequence
44	969	30.8	1788	6	AX506749 Sequence
45	969	30.8	1788	6	BD017430 Transgeni

ALIGNMENTS

RESULT 1

AX148306
LOCUS AX148306 1800 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent EP1116794.
ACCESSION AX148306
VERSION AX148306.1 GI:14347193
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1
REFERENCE
1 Iuchi, S., Kobayashi, M. and Shinozaki, K.
Transgenic plants carrying neoxanthin cleavage enzyme gene
Patent: EP 1116794-A 5 18-JUL-2001;
Riken (JP)

FEATURES

source

1..1800
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

CDS

1..1800
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41194.1"
/db_xref="GI:14347194"
/translation="MASFTATAVSGRWLGNGHTQPLSSQSSDLSYCSSLPMAARV
TRKLVNSALHTPPALHPKQSSNSPAIVKPKAGSNTKQMLFORAAALADAAEG
FLVSEKHLHPKPADPSVQIAGNFAPVNEQVRRLNFVVGKLPDSIKGVYRNGANP
LHEVTEGHHFDDGVMVHAVKFEHGSAYACFTQTNRFVQERQLGRPPPKAIGSLH
GHTGIASLMFLFYAFAAGIYDPAHGTGVANAGLYFENGRLANSEDDLPYQVQITNG
DLKTVGRFPDGOLESIMIAHPKVDPSGBLFALSYDVVSKPKVLPFRFSPDGTGKSPD
VEIOLDPTWMDHFAITNFVVPDQVFKLPEMIRGSGSPVVDKXKVARFGILDKY
AEDSNKNIDAPRCFCEHLNWAEBETDEVVVGSCMTPPDSINESDENLKSVL
EIRLNKGTSETRPILISNEDQVNLKAGVNRNMLGRKTKFAYLALAEWPVKVSGFA
KVDLTTEVKLEHLYGDNRYGEPFLFPGEGEEDGEGYILCFVHDEKTRWSELOIVNAV
SLEVEATVKLEPSRYPFGHGTFFIGADLAKQVV"

BASE COUNT 458 a 464 c 439 g 439 t

ORIGIN

Alignment Scores:

Pred. No.: 6,73e-220 Length: 1800
Score: 3150.00 Matches: 599
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-758-269-6 (1-599) x AX148306 (1-1800)

QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
DB 1 ATGGCTTCTTTCACGGCAACGCTCGGGTTTCTGGAGATGGCTTGGTGGCAATCATACT 60
QY 21 GlnProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
DB 61 CAGCGCCCAATATCGCTTCTCTCAAGCTCCGACTTGAGTTATGTAGCTCTCTTACCTATG 120
QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
DB 121 GCCAGTCGTGTCCACAGTAAAGCTCAATGTTTCATCTGGCGTTCACACTCCTCCAGCTCTT 180
QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
DB 181 CATTTTCCCCTAAGCAATCATCAAACTCTCCCGCATTTGTGTTTAAAGCCCAAGCCNAAGA 240
QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
DB 241 TCCACACATAACAGATGAATTTGTTTCCAGAGAGCGGCGCGCAGCGCTTGACCGCGCG 300
QY 101 GluGlyPheLeuValSerHisGlnLysLeuHisProLeuProLysThrAlaAspProSer 120
DB 301 GAGGTTTCTCTTGTACGACGAGAGAGCTACACCGCGCTTCTTAAACCGCTGATCCTAGT 360

121 ValGlnIleAlaGlyAsnPheAlaProValAsnGlnProValArgAsnLeuPro 140
DB GTTCAGATCGCGGAAATTTTGTCCGGTGAATGAACAGCGCCGTCGCGCTTAATCTTCOG 420
QY 141 ValValGlyLysLeuProAspSerLysGlyValTrpValArgAsnGlyAlaAsnPro 160
DB 421 GTGTGCGAAACATTCCTGATTCATCAAGGAGTGTATGTGCGCAACGAGGTAAACCA 480
QY 161 LeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaVal 180
DB 481 CTTCCAGGAGCGGTGACAGAGTCAACACTTCTTCACGAGAGCGGTATGGTTACGCGCTC 540
QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
DB 541 AAATTCGAACACGGTTAGCTAGCTAGCTTTCGCGGTTTACTCAGACTAACCGGTTGTT 600
QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
DB 601 CAGGAACGTCAATTTGGGTGCGACCGGTTTCCCAAGACCATCGGTAGCTTACCGGCAC 660
QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaAlaAlaAlaGlyIleValAspPro 240
DB 661 ACCGGTATTGCGCGACTCATGCTATTCTACCCAGAGCTGCGCGGTATAGTCGACCG 720
QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAla 260
DB 721 GCACACGGAACCGGTGTAGCTAACCGCGTTTGTCTATTTCATGGCGGTTATTGGCT 780
QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
DB 781 ATGTGCGAGGATGATTTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAC 840
QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
DB 841 GTTGTGCTGCTGATTTTGTATGGACAAATTAGAAATCCACAATGATTGCCACCCGAAAGTC 900
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
DB 901 GACCCGGAATCCGCGTGAATCTCTTCGTCTTAAAGTACGACGCTGTTTCAAGAGCTTACCTA 960
QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
DB 961 AAATATCTCCGATTTCTCCCGGACGGAATAATCACCGGAGCTCGAGATTCAGTTGAT 1020
QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360
DB 1021 CAGCCAAACGATGATGCAGATTCGCGATTACAGAGAACTTCGTCGTCTGCTACCTCACCAG 1080
QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
DB 1081 CAGTCTGTTTTCAAGCTGCGGAGATGATCCCGGTGGGTCTCCGGTGGTTTACGACAAG 1140
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnLysLys 400
DB 1141 AACAGGTGCGAAGATTCGGGATTTTAGACAAATACGCGAAGATTCATCGAACATTAAG 1200
QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
DB 1201 TGGATTGATGCTCCAGATGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGAGCCAGAA 1260
QY 421 ThrAspGluValValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
DB 1261 ACAGATGAAGTCGTCTGTATAGGTCCTGTATAGCTTCCACACAGACTCAATTTTCAACGAG 1320
QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
DB 1321 TCTGACGAGAAATCTCAAGAGTGTCTGTCTGAATCCGCTGAATCTCAAAACCGGTGAA 1380
QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
DB 1381 TCAACTCGCGCTCCGATCATCTCCACAGNAAGATCAACAAGTCAACCTCGAAGCAGGATG 1440

QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
 Db 1441 GTCACAGAAAATCATCTCGCGCGTAAACCAATTCGTTACTTGGCTTAGCCGAGCG 1500
 QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 1501 TGGCCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGAGAGAAAGTTAAGAAACAT 1560
 QY 521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGluGlyGlu 540
 Db 1561 CTTTACGGCGATTAACCGTTACGAGGAGAGAGCTCTGTCTCTCCCGAGAGAGAGAGAG 1620
 QY 541 GluAspGluGlyTyrTrpLeuLysPheValHisAspGluLysThrTrpLysSerGluLeu 560
 Db 1621 GAAGACCAAGGATACATCTCTCTGTTTCGTTTCAGCAGAGACATCGAAATCGGAGTTA 1680
 QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 1681 CAGATAGTTAGCCGTTAGCTTAGGTTGAGGACACGTTAACTTCCGTCAGAGGTT 1740
 QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
 Db 1741 CGGTACGGATTCACGCTACATTCATCGAGCGGATGATTTGGCGAAGCAGGTCGTG 1797
 RESULT 2
 BD017431
 LOCUS 1800 bp DNA linear PAT 27-AUG-2002
 DEFINITION Transgenic plant using neozanthine cleaving enzyme gene.
 ACCESSION BD017431
 VERSION BD017431.1 GI:22558607
 KEYWORDS JP 2001258579-A/3
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 1800)
 AUTHORS Iuchi, K., Kobayashi, M. and Shinozaki, K.
 TITLE Transgenic plant using neozanthine cleaving enzyme gene
 JOURNAL Patent: JP 2001258579-A 3 25-SEP-2001;
 THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
 OS Arabidopsis thaliana (thale cress)
 PN JP 2001258579-A/3
 PD 25-SEP-2001
 PF 11-JAN-2001 JP 2001003476
 PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC
 C12N15/09,A01H5/00,C12N5/10,C12N9/02,C12N15/00,C12N5/00 CC
 Transgenic plant using neozanthine cleaving enzyme gene FH
 Location/Qualifiers
 FT CDS Location/Qualifiers
 1..(1800).
 1..1800
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 458 a 464 c 439 g 439 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6 73e-220 Length: 1800
 Score: 3150.00 Matches: 599
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-758-269-6 (1-599) x BD017431 (1-1800)
 QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
 Db 1 ATGGCTTCTTTCACGGCAACGCTGCGGTTCTTGGGAGATGGCTTGGCGCAATCATACT 60
 QY 21 GlnProProLeuSerSerGlnSerAspLeuSerTyrCysSerSerLeuProMet 40

Db 61 CAGCGCCATTATTCGCTTCTCAAGCTCCGACTTCAGTTATTGTAGTCTCTTACCTATG 120
 QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
 Db 121 GCCAGTCTGTACACAGTAAGCTCAATGTTTCACTGCGCTTCCACTCTCCAGTCTT 180
 QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
 Db 181 CATTTCCCTTAAGCAATCATCAAACTCTCCGCCATTGTTGTTAAGCCCAAGCAAAGAA 240
 QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
 Db 241 TCCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCGCTTGGACGCGCGC 300
 QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
 Db 301 GAGGTTTCTTGTGAGCCACGAGAGCTACACCGCTTCTCTAAACGGGTGATCTAGT 360
 QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
 Db 361 GTTCAGATCCCGGAAATTTGCTCCGGTGAATGAACAGCCCGCTCCGGGTAAATCTTCG 420
 QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyValaAsnPro 160
 Db 421 GTGTCGGAATACTTCCGATTCCATCAAGAGAGTGTATGTGCGCAACGGAGCTAACCCA 480
 QY 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
 Db 481 CTTTCAGAGCGGTGACAGGTCACCACTTCTTCGAGCGAGACGATGTTGTTTCACGCGTC 540
 QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
 Db 541 AAATTCGAACAGGTTTCAGTACTAGCTTTCGCGTTTACTCAGACTAACCGGTTGTT 600
 QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
 Db 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGGCCATCGCTGAGCTTCACGCGCAC 660
 QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
 Db 661 ACCGTTATTCGCGACTCATGCTATTCTACGCGAGAGCTCGACCGGTATGTCGACCCG 720
 QY 241 AlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAla 260
 Db 721 GCACACGGAACCGGTGTAGCTAACCGCGTTTGGTCTATTTCATTCAGTCGCGGTTATTGGCT 780
 QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
 Db 781 ATGTCGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAACACC 840
 QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
 Db 841 GTTGGTTCGTTTCGATTTCGATTCGACCAATTAGATCCACATGATTCGCCACCCGAAAGTC 900
 QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
 Db 901 GACCCGGAATCCGTTGAACCTCTTCGTTTAAAGCTACGACGTCGTTTCAAGACCTTACCTA 960
 QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
 Db 961 AAATACTTCGATTCTCAGCGAGGAACTAAATCACCAGCGCTCGAGATTCAGTTGAT 1020
 QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
 Db 1021 CAGCCAAACGATGTCACGATTTTCGATTTACAGAACTTCGTCGTCGTCGTCGTCGTCG 1080
 QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
 Db 1081 CAAGTCGTTTCAAGCTCCCGAGATGATCCGCGTGGGTCCTCCGTTGGTTACGACAAG 1140
 QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400

```

Db      1141 AACAGGTCGCAAGATTCGGGATTTTAGACAAATACCGCAAGATTCATCGAACATTAAG 1200
Qy      401 TTPileAspAlaProAspCysPheHisLeuTTPAsnAlaTTPiGluProGlu 420
Db      1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGACGAGAA 1260
Qy      421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
Db      1261 ACAGATGAAGTCGCTGATAGGGTCCTGTATGACTCCACAGACTCAATTTTCAACGAG 1320
Qy      441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db      1321 TCTGACGAGAAATCTCAAGAGTGTCTGTCTGAATCCGCTGAATCTCAAAACCGGTGAA 1380
Qy      461 SerThrArgArgProIleLeuSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
Db      1381 TCAACTCGCCGTCGATCATCTCCACGAAGATCAACAAAGTCAACTCGAACAGGATG 1440
Qy      481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaLysLeuAlaLeuAlaGluPro 500
Db      1441 GTCAACAGAAACATGCTCGCCGTAAACCCAAATTCGTTACTTGGCTTTAGCCGAGCG 1500
Qy      501 TTPProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
Db      1501 TGGCCTAAAGTCTCAGGATTCTGCTAAAGTTGATCTCCTACTCTGAGAGTTAAGAAACAT 1560
Qy      521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlu 540
Db      1561 CTTTACGGCGATTAACCGTTACGGAGGAGAGCCCTCTGTTCTCCCGGAGAGGAGAG 1620
Qy      541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTTPlysSerGluLeu 560
Db      1621 GAAGACCAAGATACATCTCTGTTCTGTTTCACGACGAGAGACATGGAATCGGAGTTA 1680
Qy      561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db      1681 CAGATAGTTAACCGCTTACTTAGAGTTGAGCAACGGTTAACTTCGTTCAAGGGTT 1740
Qy      581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db      1741 CCGTACGGATTTCACGGTACATTCATCGAGCCGATGATTTGGCGAAGCAGGTCGTG 1797

RESULT 3
AB026549 LOCUS
DEFINITION Arabidopsis thaliana gene for neoxanthin cleavage enzyme, complete cds.
ACCESSION AB026549
VERSION AB026549.1 GI:16416373
KEYWORDS neoxanthin cleavage enzyme.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 Iuchi, S., Kobayashi, M. and Shinozaki, K.
AUTHORS Characterization of neoxanthin cleavage enzyme from Arabidopsis thaliana
TITLE Unpublished
JOURNAL 2 (bases 1 to 3869)
REFERENCE Iuchi, S. and Shinozaki, K.
AUTHORS Direct Submission
TITLE Submitted (22-APR-1999) Satoishi Iuchi, RIKEN, Plant Mol. Bio.;
JOURNAL 3-1-1 Kouyadai, Tsukuba 305-0074, Japan
(E-mail: iuchi@rtc.riken.go.jp, Tel: 81-298-36-4359)
FEATURES
location/Qualifiers
source 1..3869
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Col."
/db_xref="taxon:3702"
1831..3630

```

```

/codon_start=1
/product="neoxanthin cleavage enzyme"
/protein_id="BAB70609.1"
/db_xref="GI:16416374"
/translation="MASPTATAAASVSGRWLGGNHTPPLSSSSQSSDLSSYCSSLPMSARV
TRLNYSALHTPPALFPKSSNSPAIVVKKPKESNTKQNLFORAAAYLDAABG
FLVSHKLPKPTADPSVQIAGNPAVNEQPVRRNLVVKGLPDSIKGVYVRNAGP
LHPVTGHPFDGDMVHAKVPHGSGASVACFTQTRFVQERQLGRPVFPAIGELH
GHTGIALMLFYARAAGIVDPAHGTGVANAGLVYENGLLAMSEDLPYVOITPNG
DLKTGSEDFDGGLESTMTAHKVPDSEGLFALSVDVVKPKVLKYFRSPDGTKSD
VEIQLOQTMHDFALTENFVVDQOVVFKLPEMRGSSPVVYDKNKVAFGLDKY
AEDSSNKTADAPDFCFHNAWSEPEIDVVVIGSCMTTPDSINESDENLKSVLIS
EILNKTIGESTRPIISNEDQVNLQAGMNRNLGRKTFAYLALABPWPVKSQFA
KVLDTTGEVKKHLYGDNRYGEPFLFPGEGEEDGYTILCFVHDEKTKWSELQVNAV
SLEVEATVKLPSPVPYFGHTFGADLAKQVV"
BASE COUNT 1155 a 845 c 773 g 1096 t
ORIGIN
Alignment Scores:
Pred. No.: 1,99e-219 Length: 3869
Score: 3150.00 Matches: 599
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-758-269-6 (1-599) x AB026549 (1-3869)
Qy 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTTPLeuGlyGlyAsnHisThr 20
Db 1831 ATGGCTTCTTTCAGCGCAACGGCTGCGGTTCTGGGAGATGGCTTGGTGGCAATCATACT 1890
Qy 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 1891 CAGCGCGCATATCGTCTTCTCAAAGCTCCGACTTGAGTTGATTGACTCTCTTACCTATG 1950
Qy 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
Db 1951 GCCAGTGTGTCACACGTAAGCTCAATGTTTCACTCGCTTTCACACTCTCTCCAGCTCTT 2010
Qy 61 HisPheProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGlu 80
Db 2011 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCAATTTGTTTAAGCCCAAGCCAAAGAA 2070
Qy 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
Db 2071 TCCACACATAACAGATGAATTTGTTCCAGAGACGGCGCGCGGCGGCGGCGGCGGCGG 2130
Qy 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
Db 2131 GAGGGTTTCTTGTGTCAGCCACGAGAAGCTACACCGCTTCTCTAAACCGGCTGATCTTAGT 2190
Qy 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
Db 2191 GTTCAGATCGCGGNAATTTTGTCCGGTGTAATGACAGCCCGTCCGCGGTAATCTTCG 2250
Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db 2251 GTGGTCGGAACACTTCCCGATTCCATCAAGAGAGTGTATGTGCGCAACGGAGCTAACCCA 2310
Qy 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
Db 2311 CTTCCAGAGCCGGTGACAGGTCACCACTTCTTCCAGCGAGACGGTATGTTGTTCCAGCG 2370
Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db 2371 AAATTCGAACACGGTTCAGTACGTAGCTTGCCTTGCCTTCTTCTACAGCTAACCGGTTT 2430
Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db 2431 CAGGAACGTCATATGGTCGACCGGTTTTTCCCAAGACATCGGTGAGCTTTCACGGCCAC 2490
Qy 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240

```

Db 2491 ACCGGTATTCCCGGACTCATGCTATTCTACCGCAGAGCTGCAGCCGGTATAGTCGACCGG 2550
Qy |||||
241 AlahisgLyThrGlyValAlaAenAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
Db |||||
2551 GCACACGAACCGGTGAGCTACCGCGGTTTGGTCTATTTCATGCGCGGTATTGGCT 2610
Qy |||||
261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
Db |||||
2611 ATGTCGGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAACACC 2670
Qy |||||
281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
Db |||||
2671 GTTGTCGGTTCGATTTTGTATGACAAATTAGATCCAAATGATGCCACCCGGAAGTC 2730
Qy |||||
301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
Db |||||
2731 GACCGGAATCCGGTGAACCTTCCTGCTTTAAGCTACGACGCTGCTTTCAAAAGCCTTACCTA 2790
Qy |||||
321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
Db |||||
2791 AAATACTTCGATTTCTCACCGGACGGAATTAATCACCGGACGTCGAGATTCAGCTTGAT 2850
Qy |||||
341 GlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
Db |||||
2851 CAGCCACGATGATGCAGATTTCCGCGATTACAGAGAACTTCGTCGTCACCTGACCCAG 2910
Qy |||||
361 GlnValValPheLysLeuProGluMetIleArgGlySerProValValTyrAspLys 380
Db |||||
2911 CAAGTCGCTTTCAAGCTCCCGAGATGATCCGCGTGGTCTCCGCTGGTTTACGACAAG 2970
Qy |||||
381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db |||||
2971 AACAGGTCCGAAGATTCGGGATTTTAGACAATACCGCGAGATTCATCGACATTAAG 3030
Qy |||||
401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnIleTrpGluGluProGlu 420
Db |||||
3031 TGGATTGATGCTCCAGATGCTCTGCTGCTCCATCTCTGGAACGCTTGGGAAGAGCCAGAA 3090
Qy |||||
421 ThrAspGluValValIleGlySerCysMetThrProAspSerIlePheAsnGlu 440
Db |||||
3091 ACAGATGAGTCGTCGTATAGGTCCTGTATGATCTCCACGAGACTCAATTTTCACAGAG 3150
Qy |||||
441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db |||||
3151 TCTCAGCAGAAATCTCAAGAGTGTCTCTGTAATCCGCTGAATCTCAAAACCGGTGAA 3210
Qy |||||
461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
Db |||||
3211 TCAACTCGCTCCGATCATCTCAACGAAGATCAACAGTCAACCTCGAAGCAGGGATG 3270
Qy |||||
481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
Db |||||
3271 GTCAACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCCG 3330
Qy |||||
501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
Db |||||
3331 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTCTGAGGAAGTTAAAGAACAT 3390
Qy |||||
521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlyGlu 540
Db |||||
3391 CTTTACGGCGATACCGTTACGGAGGAGAGCTCTGTTCTCCCGGAGGAGGAGAGAG 3450
Qy |||||
541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db |||||
3451 GAAGACGAAGATACATCTCTGTTTCTGTTCCACGACGAGAGACATCGAAATCGGAGTTA 3510
Qy |||||
561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db |||||
3511 CAGATAGTTAACCGCGTTAGCTTTAGAGGTTGACCAACCGTTAACTTCGCTCAAGGTT 3570
Qy |||||
581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599

Db 3571 CCGTACGATTTCAGGTACATTTCATCGAGCCGATGATTGGCGAAGACGGTGGT 3627
RESULT 4
AB028617/c
LOCUS AB028617 52232 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone:MOA2.
ACCESSION AB028617 BA000014
VERSION AB028617.1 GI:5041970
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
Pl and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
10819329
2 (bases 1 to 52232)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (09-JUN-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan[E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934]
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MOA2
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3//),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://grmmln1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE//).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MLN21 and the 3' clone is MIE1.

FEATURES
source

1..52232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MOA2"
/clone_lib="Mitsui Pl"
Join(1996..2132,2225..2349,2455..3501,2648..2725,
2807..2931,3021..3087,3216..3278,3370..3439,3529..3629,
3718..3822,3906..4091)
/note="gene_id:MOA2.2"
/codon_start=1
/evidence=not_experimental
/product="glycolate oxidase"
/protein_id="BAB01334.1"
/db_xref="GI:11994212"

CDS	ELTWALSGCESLKEISRNHITTEWDTSPSABL" complement (join(4591..4659,4763..4924)) /note="gene_id:MOA2.3 unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BAB01335.1" /db_xref="GI:11994213" /translation="MGRRAGTRVGNRRDLSLLTRFVDSVFYFRLAEFFELFVLFM LITVFKDITSRPENRIILVXPGGSDIWF" complement (11870..13669) /note="gene_id:MOA2.4" /codon_start=1 /evidence=not experimental /product="9-cis-epoxycarotenoid dioxygenase" /protein_id="BAB01336.1" /db_xref="GI:11994214" /translation="MASFTATAVSGRWLGNTQTPLSSQSDLSYCSSLPMSRV TRKLVNSALHTPPALFPKQSSPAIVVVKPKAGSNTKQNLFORAAAALDRAAG FLVSHKHLPTKTDGMSVQIAGNFAPNEQVRRNLPPVVGKLPDSIKGYVYVNGANP LHEVTHGHPFDGDMVHAKFPHGSGASACRTQTNRVFERQRLGRPVFKAIAGELH GHTGIRMLFYPARAAAGIVDPAGHTGVANAGLVYENGRLLAMSEODLPVOQITPNG DLKTVGRDPDQGLSTMIAPKVPDPESSELFALSDVUSKPKVLYFRSPDGTSPD VEIQDQPTMHMDFAITENFVVPDQOVVKLPDMRGSGSPVVDKXVAREGILKY AEDSNNKWDAPCCFCFLHNAWEEPEDEVVVIGSCMTPTPDSIFNESDENLKSVL EIRNLKGTSTREPIISNEDQQNLEAGVNRNMLGRKTFAYLALAEWPVKVSGFA KVDLTTEGVKHYGDNRYGSGEPLFEGEGEEDGYILCFVHDEKTSKSELQIVNAV SLEVEATVKLPSRVYPYGHGTFFIGADDLAKQV" join(29980..30324,30421..30528,30616..30660,30841..30913, 31021..31133,31239..31298,31376..31492,31588..31649, 31733..31793) /note="contains similarity to RNA-binding protein gene_id:MOA2.5" /codon_start=1 /evidence=not experimental /protein_id="BAB01337.1" /db_xref="GI:11994215" /translation="NATVTEPTDVVDDVVKVSTSEKILDEGIEKSSITDSKET ESRLDKLVAMFKLMLPLAEFFPVSYPDKKNQVAKANQFLPADDPETTKQSGEE FDLDAKDNTTKRNVSGRRRLTGRIKSAQREDSIRRVYVSDIDQSTVEGLAGL FSNCGVYDCRCDGPHSVLFAFVADDOGAHEALSGLTGLGFPVRLVPSKTAI LPVNPFLFRSEDEMCRTIYCTNIDKVSQADVFNFESACGEVTRLLJGDQLH STRIAFVFLADALSALNCSGMVVSQPIRVSPSKTPVRPRITRPSTN" complement (32182..36456) /note="gene_id:MOA2.6" /codon_start=1 /evidence=not experimental /product="disease resistance complex protein" /protein_id="BAB01338.1" /db_xref="GI:11994216" /translation="MANSYLSGCANVMVERINTSQELVELCKGSSALLKRLKVALV TANPVADADORAHRVREKHLTGIDAFQADIDLELQTEALRRRVVAEGGLGG LPPNLMAGREAIQKLEPKMKVVRLEHVKHLEVLKEYSTREPQWQASRSP DLPGRVLGVEDKLANVLLSDDEISIGKPAVISVVGMPGKGTLLTEIIPNDR VTEHPKVMISAGINFNVTIKAVLDQITSSAVNTEDLPSLQIQUAKTUSKRFLL VLDDPWSFSDSEMSFOVATDAEGSKVLTTRSEIVSTVAKAEKIYQMKLMTNEEC WELISRFAGNISVGSINQELGKRIAEQCKLPKRAHAIASHLRSKNPDWYAY SKNFSYNSLIPVLKLVDPOLKPCFALCSIFPKGHVDFDEELVLLMWAIDLILY OPRSRRLEDIGNVLGDLVAGSFFORDITMTFVMDLMDLAKAVSGDFCFRLD DNIPETPTTFHPSRSQCADSVAFRSICGAERLTILPNSPTLSLQITKVLN PLNALSGRLSLSHYQITNIPSLKGLRLVLDLSSTIKELPFPVCTLQNLQTL LILSNCRDITSPKSAELINRLDLVGTPLVEMPFGKLRSLQKUSNFVIGRLSGA GLHEKLSHLGRTIRISELQNVAFSAKAGLRKFPDLGLIKTKVSGGVPVGS FNALACDQKNSLMELPHPLTKFCIESYQGGAPFKLWSDSSFFGTSVLSNCLCI SLPPYQQLPSLKLSTIEKFNILQKVLGDLFFFGENNSRGVPFQIQLIKYMPRWDEW ICPELDGIFPCLQKLIICQCPSLRKKFPEGLPSSTEVITDCPLRVSGGENFRS LTNIPESPAISPMRRELSPGKPSDASTAQOPGASSQSDNDNEVTSLSLS LPKQDQTEDPDQYEQLSLPQOQFEPAVIRVSGYISDIPSLSPKMSPTSLVLPD KNEGISLPSSSYIQHQKISVSPSSBAIKPSQDDBDETMELIKVTDISHLME LPONLQSHLSDCDGLTFENELHELLIIACHLESFSGSPPTILKILYI RCKNKLFTSPQTRYSQLEYPFGSSCNLVNFPFLSPFKRSKIDRCEBFKFL SIHAGLDRIALESIRDCNLETFPQGLPTPKLSMLLSNCKLQALPEKLFGL TSLLSLFIKCPKEIETIPGGGFPNSLRLTLCISLQDKLTPRIEWGLRDLNLRNLEIDG
CDS	GNEDIESFPBGLPKSVLSRISRFEENLKLNRKGFHDTKALETMEISQCDKLQISI DEDLPPLSCLRISCSLLTETFAEVETEPFKVLNIPYVEIDGIFS" 38132..41296 /note="gene_id:MOA2.7" /codon_start=1 /evidence=not experimental /product="disease resistance complex protein" /protein_id="BAB01339.1" /db_xref="GI:11994217" /translation="WTGIGEMFLAAFLQALFOTLVSPBPSFFPKRRELNNLLRLST ALLITAVLDAEKKQITNVEKWNELDVVYHADALDDIATEALRNILGAESS SNRLQRGRNSUGQIDTGNSEHDELEKVTIRLERLASQRNILGKELTAMIPKQR LPTTSLVDESEVFGDDDKDLMRFLIPENGKONGITVVAIVGIGVGKTLTSLGILN DQHYSYFGTKVAHVSSEDFDKVTKVYSVTSRCEFTDLDVLQKUKRLUTGIG LPFLVLDLDMNENFADLRLQFFIHAAQSQILVTRRSQVASIMCAVHVHNLQPL SDGCWSLPMKTVFGNOEPCILNREIGDLAERI VHKCRGLPLAVKTLGVLRFEGKVI WELVSRIMDLPADKNLPLVRSYVYLPAHLKRCFPAYCSIFPKGHAPEKDKVILL WMAEGFOQTRSSKNLBELNEVSELESLLQKTRVIMHDFINELAQFASGEFS SKFDECKLQVSEKTRISLYLRDNYAPMEFEALREVKFTPLFLPLTNSRSSCCLD QWSEKULPTIRLSLRVLSHYKIRLPPDFKNISHARFDDLSRLELEKLPKSLCYM YNQTLLSYCSLKEIPTDLSNLIRYLDLIGTKRQMPRRFRGLKSLGTLTTFV SASDGSRISELGHLHDHGLKIVELQVVDVADAANLNSKKHLEIDFVWRTGS SSENNTNPHRTQNEAEVPEKLRPHRIEKLAIERKGRFPDMSLDSFSRIVICRLR ECOYTSIPSGQLPCKELHISGMVGLSQISGRKFTSDOOLRDQDQPFASLTLRF DLPDMOEWLAVRTRGDLFPLAKLRLRCPELTGTLPTFPLSLISLHKCGLLDF QPDHESYRNLTLSIKSSODTLVKSPLAHFANLKLVDQCTSLYSLELSNHLRG FNAUKLNQNLQQLPLKUNALPQNLQVITNCRILRQPMEQQPYVHPHFPLPRS NVSGSPKSHGSHRSYDSSRSRYD" 42345..42872 /note="gene_id:MOA2.8 unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BAB01340.1" /db_xref="GI:11994218" /translation="MSDLVIALVAPAVLFVVFIILSCIMGEGKDVHSPPLPRPQQLR MPVYTKDILGSDGTAFISVAGTDIYGGSGVGGHGGRGITSDTKLPSPPPSPPLP PAPVORPKSVSRKTQRELIDTGTAMLVASSILLTSGSSCSGSGSHGGCGGGGGG GGLGGCGGGGGCGG" complement (join(42980..43465,43561..43668,43952..44119, 44215..44460,44539..44677,44759..44977,45060..45754)) /note="gene_id:MOA2.9" /codon_start=1 /evidence=not experimental /product="terpene synthase-like protein" /protein_id="BAB01341.1" /db_xref="GI:11994219" /translation="MKPKVRMLKSSKNSDNERIRLIELLNLGLIAYHFEIIDEILG QAFNLDDIIAKENDLETISFMFEVRLRGYMPCKSIGDLSHCILHINYLADEF VDANRFKXGDFRKESLAEDIRGMLQYEAHLGTPSEDIMDALSTFRLESLS NHTATASPLSHKIQNALYRARNLILVAREYISFYEQBEDHDTLTKLAKUNFY COLHYIQELKDLTKWMLDASKLPYRIDRIVEYFGALALYFEPYSLGRILVTKI TMIVTNDTCDAVGTLPVTSVDSFORWLDGTEKLPYKIVYFRGVFTLEEIEQ EMRPOGRSRIVOVADEIKGLKAYLAISKWARASHVTFEYMEFGMTQMDHFAAY SIAMEDCQDQTCWYKSRPMMEALANGVRIKNDINTFQEMRSGEVAKGLNCYMK QGVSKBEAIGEMNKIYNNYKIIMEEYLTAVPRPILFETDLNLSKLIIKTITLM FRLVNDVGYVEHLKTEIRGEVANGDLDFYMKQHGVTKEASQELAMKNDYKV MEEFVNDHDLPRQVFLKRNRIARFDVYITEVDGDKPKGIENFMISLYLHPINTP
CDS	Alignment Scores: Pred. No.: 7,97e-218 Length: 52232 Score: 3150.00 Matches: 599 Percent Similarity: 100.00% Conservative: 0 Best Local Similarity: 100.00% Mismatches: 0 Query Match: 100.00% Indels: 0 DB: 8 Gaps: 0 US-09-758-269-6 (1-599) x AB028617 (1-52232)
Qy	1 MetAlaserPhenThraIaThraIaAvalSerGlyVArGTriLeuGlyGlyAenHisThr 20
Db	13669 ATGGCTTCCTTACGCAACGGCTCGCGGTTCGGAGATGCGTGGTGGCAATCATCT 13610

QY 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
 Db 13609 CAGCCGCCATTTATCGCTCTCTCAAGCTCCGACTTGAGTTATGTAGCTCCTTACTATG 13550
 QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
 Db 13549 GCGAGTCGTGTACACGTAAGCTCAATGTTTCTATCTGCGCTTCACACTCCTCAGCTCTT 13490
 QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
 Db 13489 CATTTCCCTTAGCAATATCAAACTCTCCGGCATTTGTTAAGCCCAAGCCAAAGAA 13430
 QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
 Db 13429 TCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCAGCGTTGACGCGCG 13370
 QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
 Db 13369 GAGGGTTTCTGTCAGCCAGAGAGAGCTACACCGCTTCCCTAAACGGCTGATCTAGT 13310
 QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
 Db 13309 GTTCAGATCGCCGAAATTTGCTCGGTGAATGAACAGCCCGTCCGGCGTAATCTCG 13250
 QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
 Db 13249 GTGGTCGGAACCTTCCCGATTCATCAAGAGAGTGTATGCGCAACGAGCTAACCCA 13190
 QY 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
 Db 13189 CTTACAGCGCGGTGACAGCTCACCACTTCTTCGACGAGAGCGGTATGTTTACGCGCTC 13130
 QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
 Db 13129 AAATTGCAACACGCTTACGCTAGCTACGCTTCGCGTTTACTCAGACTAACCGGTTTGT 13070
 QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
 Db 13069 CAGGACGTCAATTGGGTGACCGGTTTCCCAAGAGCCATCGGTGAGCTTCACGCGCAC 13010
 QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
 Db 13009 ACCGGTATTCGCGACTCATGCTATTCTACGCGCAGAGCTGCAGCGGTATAGTCGACCG 12950
 QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
 Db 12949 GCACACGAACCGGTGTAGCTAACCGCGTTTGGTCTATTCTAATGCCGGTTATGGCT 12890
 QY 261 MetSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
 Db 12889 ATGTGCGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACC 12830
 QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
 Db 12829 GTTGGTCGGTTCCGATTTTGTATGACAAATTAGATTCACAATGATTCGCCACCCGAAAGTC 12770
 QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
 Db 12769 GACCCGGAATCCCGTGAACCTTCGCTTTAAGCTACAGCTGCTTTCAAAGCCTTACCTA 12710
 QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
 Db 12709 AAATCTTCGATTTCTCACCGGACGGAATTAATACCGGACGTCGAGATTCAGCTTGTAT 12650
 QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360
 Db 12649 CAGCCACGATGATGCAGATTTCCGATTTACAGAGAACTTCGTGCTGCTACCTGACCG 12590
 QY 361 GlnValValPheLysLeuProGluMetIleArgGlySerProValValTyrAspLys 380
 Db 12589 CAAGTCGTTTCAAGCTCCCGAGATGATCCGCGGTGGTCTCCGGTGTTCAGCAAG 12530
 QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400

Db 12529 AACAAAGTCCAAAGATTCGGGATTTTAGACAAATACGCCAAGATTCATCGAATTAAG 12470
 QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
 Db 12469 TGGATTGATCTCAGATTCT 12410
 QY 421 ThrAspGluValValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
 Db 12409 ACAGATCAAGTCGTGCTGATAGGTCTCTGATGACTCCACAGACTCAATTTTCAACGAG 12350
 QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 Db 12349 TCTACAGAGATCTCAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12290
 QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
 Db 12289 TCAACTCGCGTCCGATCATCTCCACAGAGATCAACAGCTCAACCTCGAAGCAGGATG 12230
 QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrIleAlaLeuAlaGluPro 500
 Db 12229 GTCAACAGAAACATGCTCGGCCCTTAAACCAATTCGCTTACTTGGCTTTAGCGGACCG 12170
 QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 12169 TGGCTTAAGTCTCAGATTTCGCTAAGTTGATCTCACTACTGGAGAAGTTAAGAAACAT 12110
 QY 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlyGlu 540
 Db 12109 CTTTACGCGATACCGCTTACGAGGAGAGCTCTGTTTCTCCCGGAGAGGAGAGAG 12050
 QY 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
 Db 12049 GAAGACGAGGATACATCT 11990
 QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 11989 CAGATAGTTAACGCGGTAGCTTAGAGTTGAACAAACGGTTAACTTCCTCAAGGTT 11930
 QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
 Db 11929 CCGTACGATTTACGGTACATTCATCGAGCCGATGTTTGGCAAGCAGCTGCTG 11873

RESULT 5

AY056255

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A. Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

Riken Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN Arabidopsis Full-length cDNA"): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Kossena, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis Genome submitted to GenBank.

Location/Qualifiers

1. .2331

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAFL08-11-H16 (R11094)"

/note="This clone is in a modified pBluescript vector (FLC-1) as a BamHI/XhoI insert."

ecotype: Columbia

1. .2331

/gene="At3g14440"

1. .122

/gene="At3g14440"

123. .1922

/gene="At3g14440"

/codon_start=1

/evidence="experimental"

/product="putative 9-cis-epoxycarotenoid dioxygenase"

/protein_id="RAFL07104.1"

/db_xref="GI:15810433"

/translation="WASETATAAVSGRWLGNNHTQPPPLSSQSSDLSCSLPMSRVTRKLVSSALHTPPALHPFKOSSNFAIVVFKAKESNTKQNNLFQRAAALDRAEGFLVSEKHLPLPKTADPSVQIAGNFPAPYNEQPVRRNLFPVVGKLPDSIKGVYRNGANPLHEPVTGHFFDGDGMHAKVFEHGSASYACRFTQTNRFVQERQLGRFPVFKAIHELHGTGLARMLFYARAAAGIVDPAHGTGVANAGLVFNGLRLAMSEDDLPYQVQITPNGLDKTVGRDPDGLSMTIAHPKVPESGELFALSYDVVSKPKLYKPRFSPDQKSPDVLQDQPMHDFALTENFVVPDQVVFVKLPKEMILGSPVVDVKNKVARFGILDKYADSSNKIKWDFPCFCHLWNANWEPSTDEVVIGSCMTFPDSIFNEDSNLXSVLEIRLNKIGESFTRPIIISNDDQNLKAGMVRNMLRKTKFAYLALAPKPVKSGFAKYDLATTEGVKHLVYDGNRYGEPFLPQEGGEGEGYILCFVHDEKTKWSELQIVNAVSLVEVATVKLPSPVYPGHGTPIGADLAKQVV"

misc_difference 1234

/gene="At3g14440"

/note="compared to genomic sequence resulting in an amino acid sequence difference"

/replaces="g"

1923. .2331

/gene="At3g14440"

misc_difference 2059

/gene="At3g14440"

/note="not present in genomic sequence"

misc_difference 2312

/gene="At3g14440"

/note="not present in genomic sequence"

misc_difference 2315

/gene="At3g14440"

/note="not present in genomic sequence"

misc_difference 2315

/gene="At3g14440"

/note="not present in genomic sequence"

BASE COUNT 615 a 560 c 533 g 623 t

ORIGIN

Alignment Scores: 3,14e-219 Length: 2331

Score: 3143.00 Matches: 598

Percent Similarity: 99.83% Conservative: 0

Best Local Similarity: 99.83% Mismatches: 1

Query Match: 99.78% Indels: 0

DB: 8 Gaps: 0

US-09-758-269-6 (1-599) x AY056255 (1-2331)

QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTTPLeuGlyGlyAsnHisThr 20

Db 123 ATGGCTTCTTTCACGGCAACGGCTGCGGTTCTTGGAGATGGCTGGGGCAATCACTACT 182

QY 21 GlnProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40

Db 183 CAGCGCCATTATCGTCTTCTCAAGCTCCGACTGAGTTATTTAGCTCTTACCTATG 242

QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60

Db 243 GCAGTCGTGTACACGTAAGCTCAATGTTTCATCTGCGCTTCACACTCTCCAGCTCTT 302

QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80

Db 303 CATTTTCCCTTAGCAATCATCAAACTCTCCGCCATTGTTTAAAGCCCAAGCCAAAGAA 362

QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100

Db 363 TCCACACTAAACACAGATGAATTTGTTCCAGAGAGCGGGCGGACGCTTGACCGCGCG 422

QY 101 GluGlyPheLeuValSerHisGlyLysLeuHisProLeuProLysThrAlaAspProSer 120

Db 423 GAGGGTTTCTTGTTCAGCCAGAGAGCTACACCCGCTTCTTAAACCGGCTGATCTAGT 482

QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140

Db 483 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCGCTCGGCGTAACTTCGG 542

QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160

Db 543 GTGGTCGGAATACTTCCCGATTCCCATCAAGAGAGTGTATGTGCGGACACGAGCTAACCA 602

QY 161 LeuHisGluProValThrGlyHisHisPheAspGlyAspGlyMetValHisAlaVal 180

Db 603 CTTTCAGAGCGCGGTGACAGGTCAACACTTCTTTCGACGGAGACGGTATGTTTCAACGCGTC 662

QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200

Db 663 AAATTGGAACACGGTTTCAGTAGCTACGTTGCGGTTTACTCAGACTAACCGGTTTGT 722

QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220

Db 723 CAGAACGTCATTTGGTTCGACCGGTTTTCACCAAGCCATCGGTGAGCTTCAACGCCAC 782

QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaAlaValAspPro 240

Db 783 ACCGGTATTCGCCGACTCATGCTATTCTACGCGAGAGTGCAGCCGGTATGTCGACCGCG 842

QY 241 AlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260

Db 843 GCACACGAACCGGTGTAGCTAACGCGGGTTTGGTCTATTTCATGCGCGTTATTGGCT 902

QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280

Db 903 ATGTGCGGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGAGATTTAAAAACC 962

QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300

Db 963 GTTGGTCGGTTCCGATTTTGTGACACATATAGATCCACATGATTTGCCACCCGGAAGTC 1022

RESULT 8
LES439079 19018 bp DNA linear PLN 13-MAR-2003
LOCUS
DEFINITION Lycopersicon esculentum nced1 gene for 9-cis-epoxycarotenoid
dioxxygenase.
ACCESSION
VERSION A0439079
KEYWORDS 9-cis-epoxycarotenoid dioxxygenase; nced1 gene.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE
AUTHORS Thompson A.J.
TITLE Completion of notabilis, a tomato mutant deficient in the
ascibic acid biosynthetic enzyme 9-cis-epoxycarotenoid dioxxygenase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19018)
AUTHORS Thompson A.J.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2002) Thompson A.J., Plant Genetics and
Biotechnology, Horticulture Research International, Wellesbourne,
Warwick, CV35 9EF, UNITED KINGDOM
REMARK revised by author [14-MAY-2002]
COMMENT Related mRNA entry Z97215.
FEATURES
source
1..19018
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultiivar="Money-maker"
/db_xref="taxon:4081"
/chromosome="7"
/map="7q40"
/clone="NCEd1.1"
267..567
/rpt_family="tomato anionic peroxidase inverted repeat
(TAPTR)"
repeat_region
/rpt_type=INVERTED
2343..3038
/rpt_type=DIRECT
5304..7570
/gene="nced1"
5304..5311
/gene="nced1"
5334..5337
/gene="nced1"
5404..7570
/gene="nced1"
5482..7299
/gene="nced1"
/function="ascibic acid biosynthesis"
/codon_start=1
/product="9-cis-epoxycarotenoid dioxxygenase"
/protein_id="CAD30202.1"
/db_xref="GI:28974077"
/translation="MATTTSHTATNTWIKLSPSSKEFGFASNSISLLKNQHNROSL
NINSLQAPLILHPKQSNVOTPKNTTISHPKQNNNSSSSTSKWNLVQKAAAMAL
DAVESALTKHELHPKLTADPRVOISGNFAPVPENPVCSLPTVTKPKVCQGVYR
NGANPLFPTAGHHEFDGDMVHVOFKNGSVACRFTETELVQSKALGRPVFKA
IGELGHGIAKLMLFARGLFGLVHDSKGVANAGLVYFNRLLANSEDDLLPHVK
VPTGDLTEFGDFDGLKSTMAHPKLPDVGELFALSVDVQPKLYKTRFKNG
EKNSDVEIPVEDTWHDFATENFVVPDQVFPKSEMIRGGSPVYDKNKVRFG
ILDYKDGSDLKWEVDFPCFHLNNAWEAEATDEIVVIGSCMTFDFINECDEGL
KVLSEIRLNLTKGSTRKSIENDEQVNLKAGMVRNKLGRKTEYAYLAIAEPFK
VSGFAKNLFTCEVEKFTYGDNKYGGELFLPRDPNSKEEDDGVILAFVHDEKEWKE
LQIVNMSLKLEATYKLSRPVYPYGHGTFINLANDLANQA"
7570
/gene="nced1"
/evidence=experimental
8285..8575
/rpt_family="tomato anionic peroxidase inverted repeat
(TAPTR)"
repeat_region

repeat_region
10047..10285
/note="putative"
/transposon="stowaway element"
10047..10063
/rpt_type=INVERTED
/rpt_type=TERMINAL
10359..10756
/note="putative mobile element"
BASE COUNT 6823 a 2631 c 2645 g 6919 t
ORIGIN
Alignment Scores:
Pred. No.: 2.48e-155 Length: 19018
Score: 2282.50 Matches: 436
Percent Similarity: 81.83% Conservative: 64
Best Local Similarity: 71.36% Mismatches: 92
Query Match: 72.46% Indels: 19
DB: 8 Gaps: 6
US-09-758-269-6 (1-599) x LES439079 (1-19018)
QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
Db 5476 GTAGTATGGCAACTACTTTCATCGCACAAATACATCGATT----- 5520
QY 21 GlnProProLeuSerSerGlnSerSerAspLeuSerTyrCysSer-----Ser 37
Db 5521 AAGACTAAGTTTGTCAATGCCATCATCAAGGAGTTTGGTTTTCATCAAACTCTATTTC 5580
QY 38 LeuProMetAlaSerArgValThrArgLysLeuValSerSerAlaLeuHisThrPro 57
Db 5581 CTACTCAAAATCAACATAATAGGCAAGTCTCAACATTAATTCCTCTCTCAAGCTCCA 5640
QY 58 ProAlaLeuHisPheProLysGlnSerSerAsn-----SerProAla-----lleVal 73
Db 5641 CCTACTCTTCAATTTCTTAAACAATCTTCAAAATTAACAACACCAAGATAATAACAAT 5700
QY 74 ValLysProLysAlaLysGlnSerAsn-----ThrLysGlnMetAsnLeu 88
Db 5701 TCACACCCAAAACAGAAAACCAACACTCTCTCTTCTTCAACTTCCAAGTGGGATTTA 5760
QY 89 PheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108
Db 5761 GTGCAGAAAGCAGCAGCAATGCTTTAGATGCTGTAGAAAGTCTTTAACTAAACATGAA 5820
QY 109 LysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128
Db 5821 CTTGAACACCCCTTTGCCGAAAACAGCCGACCCACGATTCAGATTCTTGGGATTTTGGT 5880
QY 129 ProValAsnGluGlnProValArgAsnLeuProValValGlyLysLeuProAspSer 148
Db 5881 CCGGTACCGGAAAATCCAGTCTGTCAATCTCTTCGGGTCAACGAAAAATACCCAAATGT 5940
QY 149 IleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168
Db 5941 GTTCAAGGCGTTTACGTTTCGAACGAGCTAACCTCTCTTTTGAACCAACCCCGGACAC 6000
QY 169 HisPhePheAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188
Db 6001 CATTTCTCGACGCGACGGTATGTTTACGCGCGTTCATTTCAAAAATGGTGGCTAGT 6060
QY 189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgPro 208
Db 6061 TAGCGTTGCGGTTTCTACTGAAACAGAGAGCTTGTTCAGAAAAAAGCTTTGGTGCCTCT 6120
QY 209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228
Db 6121 GTTTTCCCTAAACCCATTGGTGAATTCATGGTCACTCTGGAATTCGAAGGCTTATGCTG 6180
QY 229 PheTyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248
Db 6181 TTTTACGCTCGTGGGCTCTTCGGACTTGTGTGATCAGACAGTAAAGAACTGGTGTGCAAC 6240

```
QY 249 AlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyr 268
DQ 6241 GCCGGTTAGTCTATTTCATTAACCCGATTACTTGTCTATGCTGAGATGATTGCTTAC 6300
QY 269 GlnValGlnIleThrProAsnGlyAspLeuLeuValThrValGlyArgPheAspGly 288
DQ 6301 CATGTAAGGTAAACACCCCGCGATCTTAAACACAGAGGTCGATTTCGATTCGACGCG 6360
QY 289 GlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPhe 308
DQ 6361 CAGCTAAATCCACCATGATAGTCTACCCAAAGCTCGACCCGATTTCCCGTGAGCTATT 6420
QY 309 AlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAsp 328
DQ 6421 GCTCTTAGCTACGATGATGATTCAGAGCCATACCTCAAGTACTCTCAGATTTTCAAAAAT 6480
QY 329 GlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPhe 348
DQ 6481 GGGGAAATAATCAAAATGATGTTGAAATTCAGATTGAAGACCCCAACATGATGATGATTC 6540
QY 349 AlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGlu 368
DQ 6541 GCATTAATGAGAACTTCGTCGTCATCTCTGATCAACAGTCGTTTCAAGATGCTGAA 6600
QY 369 MetIleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIle 388
DQ 6601 ATGATCCGTGGAGGTTCCACCGTGGTTTACGACAAAGCAAAAGTTTCCCGATTGCTATT 6660
QY 389 LeuAspLysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPhe 408
DQ 6661 CTGGTAAGTACCGGAAAGATGGTCTGATTTGAAATGGTGAAGTACTGATTTGTTTC 6720
QY 409 CysPheHisLeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGly 428
DQ 6721 TGTTCCACCTCTGGAATGCTTGGGAAAGACGAGAAACAGATGAAATCGTTGTAATTGT 6780
QY 429 SerCysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerVal 448
DQ 6781 TCATGTATGACACCCAGACTCCATTTTCAATGAATGTGATGAGGCTTAAGATGTT 6840
QY 449 LeuSerGluIleArgLeuAsnLeuLysThrGlyLysSerThrArgArgProIleIleSer 468
DQ 6841 TTATCCGAAATCCGTCTCAATTTGAAACAGGGAATCAACAGAAATCCATAATCGAA 6900
QY 469 AsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArg 488
DQ 6901 AACCCGATGAACAAAGTGAATTTAGAGCTGGAATGTTGAAACCGAACAACATCGGAAG 6960
QY 489 LysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAla 508
DQ 6961 AAAACAGAGTATGCTTATTTGGCTATCGCTGAACCATGGCCAAAGTTTCTGTTTTCGA 7020
QY 509 LysValAspLeuThrThrGlyGluValLysHisLeuTyrGlyAspAsnArgTyrGly 528
DQ 7021 AAAAGTAAACCTGTTCCACCGTGAAGTTGAGAAATTCATTTATGTTGACAAATATGTT 7080
QY 529 GlyGluProLeuPheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIle 546
DQ 7081 GGGGAACTCTTTTTTACCAGAGACCCCAACAGACGAGGAGGACGATGGTTATATT 7140
QY 547 LeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaVal 566
DQ 7141 TTAGCTTTCGTTTCAGATGAGAAAGATGGAATTCAGAACTGCAAAATGTTTAACGCAATG 7200
QY 567 SerLeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGly 586
DQ 7201 AGTTTGAAGTTGGAGGCACTGCTGAGCTTCCATCAAGAGTTCCTTATGATTTTCATGGA 7260
QY 587 ThrPheIleGlyAlaAspAspLeuAlaLysGln 597
DQ 7261 ACATTCATAAAGCGCAATGATTGGCAATCAG 7293
```

```
RESULT 9
AX148316
LOCUS 1818 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 15 from Patent EP116794.
ACCESSION AX148316
VERSION AX148316.1 GI:14347203
KEYWORDS
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; Lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
1
REFERENCE
1 Iuchi,S., Kobayashi,M. and Shinozaki,K.
Transgenic plants carrying neoxanthin cleavage enzyme gene
Patent: Ep 1116794-A 15 18-JUL-2001;
Riken (JP)
FEATURES
source
1..1818
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/db_xref="taxon:4081"
1..1818
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41199.1"
/db_xref="GI:14347204"
/translation="MATTSHANTWTKLSPSSKEFGFASNSILLKNQNRQSL
NINSLQAPILHPFKOSSNYQTPKNTIISHPKENNNSSSSTKWNLYVQKAAVAL
DAVSAALTKELEHPKTPKAPVQISGNFAPVPCQSLPVTGIPKRCQGVYVR
NGAMLPETPAHHFFDQGMVAVQPKNGSAYACRTETIRLVQEKALGRPVYVKA
IGELHSGIARLMLEFVARGLGLVDHSGKTGVANAGLVVFNRLLAMSDLLPYPVK
VTPTGDLTKSRGPDFDGLSTMTAHKLPDVSGLFALSDVIQKPLKYFRFSKNG
EKNDVIPEDPTMGADPAITENFVVIPOQVVFKNSEMRGSGSPVVDKPKYSRG
ILDYAKQSDGLKWVEPDFCFHLMNABEETDELVIGSCMTPDPSIFNEDEGL
KSVLSERLMLTKGTRKSIENPDQVNLBGMVNRNKLGRTEYAILAIAEPWPK
VSGFAKVLFTGEVKEFIYGDNGKGEPLFLPRDPSKEDDDGVLAVHDEKWKSE
LQIVNMSLKLEATVKLPSRPVYGFHGTFFNANDLANQA"
BASE COUNT 569 a 383 c 369 g 497 t
ORIGIN
Alignment Scores:
1.24e-156 Length: 1818
Pred. No.: 2280.50 Matches: 435
Score: 82.04% Conservative: 63
Percent Similarity: 71.66% Mismatches: 90
Best Local Similarity: 72.40% Indels: 19
Query Match: 6 Gaps: 6
DB:
US-09-758-269-6 (1-599) x AX148316 (1-1818)
QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24
DQ 7 ACTACTACTTCCATGCCCAAAATACATGGATT-----AAGACTAAGTTG 51
QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41
DQ 52 TCAATGCCCATCATCAAGGAGTTGGTTTTCATCAAACTCTATTCTCTACTCAAAAT 111
QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHis 61
DQ 112 CAACTAATAGGCAAGTCTCAACATTAAATCTCTCTCTTCAAGCTCCACCTATATTCTAT 171
QY 62 PheProLysGlnSerSerAsn-----SerProLala-----IleValValLysProLys 77
DQ 172 TTTCTTCAACCAATCTTCAAAATTTATCAACACACCAAGGATAATAACAATTTCAACCCAAA 231
QY 78 AlaLysGluSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92
DQ 232 CAAGAAACAAACAACTCTCTCTTCTTCAACTTCCAAAGTGAATTTAGTCGAGAAAGCA 291
QY 93 AlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112
```

292 GCAGCAATGGCTTAGATGCTGTAGAAAGTGGCTTTAACTAAACATGAACCTTGAAACACCT 351
113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132
352 TTGCGAAACAGCCAGCCACGAGTTCCTGGGAATTTCTCCGGTACCGGAA 411
133 GlnProValArgAsnLeuProValValGlyLeuProAspSerIleLysGlyVal 152
412 AATCCAGTCTGTAATCTCTCCGGTACCGGAAATATCCCAATGTGTCAAGCGTT 471
153 TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPheAsp 172
472 TAGCTTCGAACGAGCTAACCCTCTTTTGAACCAACCGCGACACCACTTTCTTCGAC 531
173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerValAlaCysArg 192
532 GGCACCGGTATGGTTCACCGCTTCAATTCARAAATGGTGGCTATGATTCAGCTTCCCGT 591
193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212
592 TTCACTGAAACAGAGAGCTTGTTCAGAAAAAGCTTTGGGTGCGCTGTTTCCCTAAA 651
213 AlaIleGlyLeuLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArg 232
652 GCCATTGGTGAATACATGGTCACTCTGGAATTCGAAGGCTATGCTGTTTTACGCTCGT 711
233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaGlyLeuVal 252
712 GGGCTCTCGGACTGTGTTGATCACAGTAAAGAACTGGTGTGCAACCGCGGTTAGTC 771
253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnValGlnIle 272
772 TATTTCATTAACCGATTAATCTGCTATGCTCAAGTACTTCAGATTTGCCCTTACCATGTAAGGTA 831
273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheGlyGlnLeuGluSer 292
832 ACACCCACCGGGATCTTAAACAGAGGGTGCATTCGATTTCCGACGCCAGCTAAATATCC 891
293 ThrMetIleAlaHisProLysValAspProGluSerGlyLeuLeuPheAlaLeuSerTyr 312
892 ACCATGATAGTCAACCAAGCTCGACCACTTCCGGTGGAGTATTTCCTTTAGTCTAC 951
313 AspValValSerLysProTyrLeuLysThrPheArgPheSerProAspGlyThrLysSer 332
952 GATGTGATTCAGAACCTACTCAAGTACTTCAGATTTCAAAAAATGGGAAAAATCA 1011
333 ProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGlu 352
1012 AATGATGTTGAAATTCAGATTCAGACCCCAACAATGATGATGATTTCCGAATTAATCTAG 1071
353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372
1072 AACTTCGTCCTGCTCCTGATCAACAGTCTGTTTCAAGATGCTGTAATGATCCGTGGA 1131
373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392
1132 GGTTCACCGGTGGTTTACGACAAAGAAAGTTTCCCGATTTGGTATTCTCGGATAAGTAC 1191
393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeu 412
1192 GCGAAAGATGGGTCTGATTTGAATGGTGGTTGAAGTACCTGATGTTCTGTTTCCACCTC 1251
413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432
1252 TGGATGCTTGGGAAAGACAGAAACAGATGAAATGCTGTGTAATTTGGTTTCATGATGACA 1311
433 ProProAsnSerIlePheAsnGluSerAspGluAsnLeuLysSerValIleSerGluIle 452
1312 CCACCAAGATCCATTTTCAATGAATGTGATGAAGGGCTAAGAGGTGTTTATTCGGAATC 1371
453 ArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGln 472
1372 CGTCTCAATTTGAACAGGGAATCAACAGAAATCCATTAATCGAAACCCGGATGAA 1431

QY 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492
Db 1432 CAACTGAAATTTAGAACTGGATGGTGAACCGAAACAACTCGGAAGAAACAGAGATAT 1491
QY 493 AlaTyrLeuAlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAspLeu 512
Db 1492 GCTTATTTGGCTATCGCTGAACCATGGCCAAAAGTTTCTGGTTTGCAAAAGTAAACCTG 1551
QY 513 ThrThrGlyValValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532
Db 1552 TTCACCGGTGAAGTTGAAATTCATTTATGGTGACAAATATATGTTGGGAACCTCTT 1611
QY 533 PheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIleLeuCysPheVal 550
Db 1612 TTTTACCAAGAGACCCCAACAGCAAGAAAGACGATGCTTATATTTAGCTTTTCGTT 1671
QY 551 HisAspGlyLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570
Db 1672 CACGATGAAAGAAATGGAATCAGAACTGCAAAATTTGTAACGCAATGAGTTTGAAGTTG 1731
QY 571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590
Db 1732 GAGGCAACTGTGAAGCTTCCATCAAGATTCCTTATGATTTTCATGGAACATTCATAAC 1791
QY 591 AlaAspAspLeuAlaLysGln 597
Db 1792 GCCAATGATTTGGCAATCAG 1812

RESULT 10

BD017436 1818 bp DNA linear PAT 27-AUG-2002
LOCUS Transgenic plant using neozanthine cleaving enzyme gene.
DEFINITION
ACCESSION BD017436
VERSION BD017436.1 GI:22558612
KEYWORDS JP 2001258579-A/8.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE 1 (bases 1 to 1818)
AUTHORS Iuchi,K., Kobayashi,M. and Shinozaki,K.
TITLE Transgenic plant using neozanthine cleaving enzyme gene
JOURNAL Patent: JP 2001258579-A 8 25-SEP-2001;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Lycopersicon esculentum (tomato)
PN JP 2001258579-A/8
PD 25-SEP-2001
PF 11-JAN-2001 JP 2001003476
PI KIYOSHI IUCHI,MASATOMO KOBAYASHI,KAZUO SHINOZAKI PC
C12N15/09,A01H5/00,C12N5/10,C12N9/02,C12N15/00,C12N5/00 CC
Transgenic plant using neozanthine cleaving enzyme gene PH
Key

FEATURES

source (1). (1818).
Location/Qualifiers
1. .1818
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/db_xref="taxon:4081"
BASE COUNT 569 a 369 g 497 t
ORIGIN

Alignment Scores:

Pred. No.: 1,24e-156 Length: 1818
Score: 2280.50 Matches: 435
Percent Similarity: 82.04% Conservative: 63
Best Local Similarity: 71.66% Mismatches: 90
Query Match: 72.40% Indels: 19
Gaps: 6
DB:

US-09-758-269-6 (1-599) x BD017436 (1-1818)

QY 5 ThrAlaThrAlaAlaValSerGlyArgTTPLeuGlyGlyAsnHisThrGlnProProLeu 24
Db 7 ACTACTACTTCATGCCACAAATACATGATT-----AAGACTAAGTTG 51
QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41
Db 52 TCAATGCCATCATCAAGAGGTTGGTTTGGATCAAACTCTATTTCTCTACTCAAAAT 111
QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeuHis 61
Db 112 CACATATAAGGCAAGCTCTCAACATTAATCTCTCTTCAAGCTCCACCTACTTCAAT 171
QY 62 PheProLysGlnSerSerAsn-----SerProAla-----IleValValLysProLys 77
Db 172 TTTCTCTAAACAATCTTCAAAATTAACAACACCAAGAAATAATAAATTTTCACACCCAAA 231
QY 78 AlalysGlnSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92
Db 232 CAGAAACACACACTCTCTTCTTCACTTCCAGTGGATTTAGTCGAGAAAGCA 291
QY 93 AlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112
Db 292 GCAGCAATGCTTTAGATGCTGTAGAAAGTGTCTTTAACTAAACATGAACCTTGAAACCCCT 351
QY 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132
Db 352 TTGCGGAAACAGCCGACCCAGAGTCCAGATTTCTGGGAATTTGCTCGGTACCGGA 411
QY 133 GlnProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal 152
Db 412 AATCCAGTGTCTCAATCTTTCGGTCCCGGAAAAATACCAAAATGTGTTCAGGCGTT 471
QY 153 TyrValArgGlnGlyAlaAsnProLeuHisGlnProValThrGlyHisPhePheAsp 172
Db 472 TACGTTCCGAACGGAGTAAACCTCTTTTGAACCAACCCCGGACACCATTTCTTCGAC 531
QY 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg 192
Db 532 GGCAGCGTATGTTTCAGCGGTTCAATTCAAAAATGGTGGCTAGTTACGCTTCCGT 591
QY 193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212
Db 592 TTCTACTGAAACAGAGAGGCTTGTTCAGAAAAAGCTTTGGGTGCGCTGTTTTCCTTAA 651
QY 213 AlaIleGlyGluLeuHisGlyHisThrGlyLeAlaAlaArgLeuMetLeuPheTyrAlaArg 232
Db 652 GCCATTGGTGAATTAATGCTCACTCTGGAATTCAGAGGCTTATGCTGTTTTACGCTCGT 711
QY 233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaAlaGlyLeuVal 252
Db 712 GGGCTCTTCGGACTTGTTCATCAGTAAAGGAACCTGTTGCAACGCGCGTTTAGTC 771
QY 253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272
Db 772 TATTTCAATACCGATTACTTGTCTATGCTGAAGATGATTGCTTACCATGTAAGGTA 831
QY 273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292
Db 832 ACACCCACCGCGATCTTAAACACAGAGGCTGATTTCGATTTTCGACGCCAGCTAAATCC 891
QY 293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312
Db 892 ACCATGATAGTCTACCCCAAGCTCGACCCAGTTTCGGGTGAGCTATTGCTCTAGTAC 951
QY 313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332
Db 952 GATGTGATTCAGAGCCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCA 1011
QY 333 ProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGlu 352
Db 1012 AATGATGTTGAATTCAGATGAGACCAACATGATGCAATGATTTCGCAATTTACTGAG 1071
QY 353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372

Db 1072 AACTTCGTCGTCATCTCTGATCAACAAGTGTGTTTCAAGATGCTGAAATGATCCCGGA 1131
QY 373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392
Db 1132 GGTTCACCGTGGTTTACGACAGAACAAAGTTTCCCGATTTCGGATTCTGGATAAGTAC 1191
QY 393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeu 412
Db 1192 CGGAAGATGGTCTGATTGAAATGGTTGAAGTACCTGATTGTTTCTGTTCCACCTC 1251
QY 413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432
Db 1252 TGGATGCTTGGGAAGACGACAAACAGATCAATCGTTGTAATGTTCTCATGTATGACA 1311
QY 433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIle 452
Db 1312 CCACGACACTCCATTTTCAATGATGATGAAGGCTTAAAGAGTGTGTTTATCCGAATC 1371
QY 453 ArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGln 472
Db 1372 CGTCTCAATTTGAACAGGGAATCAACAGAAAAATCCATATCGAAAAACCCGATGAA 1431
QY 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492
Db 1432 CAAGTGAATTTAGAAGCTGGAATGCTGAACCGAAACAACTCGGAAGAAAAACAGATAT 1491
QY 493 AlaTyrIleuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512
Db 1492 GCTTATTTGGCTATCGCTGAACCATGGCCAAAAGTTTCTGGTTTTGCAAAAGTAAACCTG 1551
QY 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeu 532
Db 1552 TTCACCGTGAAGTTGAGAAATTCATTTATGTTGACAACAATATGTTGGGGAACCTCTT 1611
QY 533 PheLeuProGlyGlu-----GlyGlyGluLysGluGlyTyrIleLeuCysPheVal 550
Db 1612 TTTTACCAGAGAGACCCCAACAGCAAGGAAGACGATGTTATATTTAGCTTTCTGTT 1671
QY 551 HisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570
Db 1672 CACGATGAGAAAGATGGAATCAGAACTGCAAAATGTTAAACGAATGAGTTGAAGTTG 1731
QY 571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590
Db 1732 GAGGCAACTGTGAAGCTTCCATCAAGATTCCTTATGGAATTCATGGAACATTCAATAAC 1791
QY 591 AlaAspAspLeuAlaLysGln 597
Db 1792 GCCAATGTTGGCAATCAG 1812

RESULT 11

AF190462

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

1 (bases 1 to 2398)

Qin, X. and Zeevaert, J. A.

The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory

step of abscisic acid biosynthesis in water-stressed bean

Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (1999)

20079657

10611386

2 (bases 1 to 2398)

2398 bp mRNA linear PLN 19-JAN-2000
Phaseolus vulgaris 9-cis-epoxycarotenoid dioxygenase (NCED1) mRNA,
complete cds.

AF190462.1 GI:6715256

Phaseolus vulgaris

Phaseolus vulgaris

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Phaseolus.

1 (bases 1 to 2398)

Qin, X. and Zeevaert, J. A.

The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory

step of abscisic acid biosynthesis in water-stressed bean

Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (1999)

20079657

10611386

2 (bases 1 to 2398)

AUTHORS Qin, X. and Zeevaert, J.A.D.
 TITLE Direct Submission
 JOURNAL Submitted (27-SEP-1999) MSU-DOE Plant Research Laboratory, Michigan State University, Wilson Str., East Lansing, MI 48824, USA
 FEATURES Location/Qualifiers
 source 1..2398
 /organism="Phaseolus vulgaris"
 /mol_type="mRNA"
 /cultivar="Top Crop"
 /db_xref="taxon:3885"
 gene 1..2398
 /gene="NCED1"
 CDS 124..1971
 /gene="NCED1"
 /function="catalyzes the oxidative cleavage of the C-11,12 double bond of 9-cis-epoxycarotenoids (C40) to yield xanthoxin (C15) and C25-apo-aldehydes"
 /codon_start=1
 /product="9-cis-epoxycarotenoid dioxygenase"
 /protein_id="AAF26356.1"
 /db_xref="GI:6715257"
 /translation="MPSPASNTWINTTLPSQSPFKDLASTSSPTLLPKKRSSS NTNITCSQTLHPKOYQFISSTTTTPPIPKPTADPKVQIAGNPAFVPEHAADQALPV QKMFLOKAATGDMVETALVSHESKHPLEKTPADPKVQIAGNPAFVPEHAADQALPV VQKPKCIDGVYVANGANPLYEPVAGHFFDGDGMHAKFTNGASVACRFETETQRL AQEKSLGRVPFKAGELHSGIARLLFYARSLFQLVDSHGMGVANAGLYTFNNH LLMSDEDLPYHVAITSGNGLTTVGRYDFNGQNLSTIAHPKLDPVNGDLHALSYDVV QKPLYRFESADGVKSPDVEIPLKEPTWMDPAITENFVVPDQVFPKLEMTITGG SPVYDNKNRSRGIILDKANAMWIDAPCFPLANAMEEPTDEIVIGSCM TPASIFNECBESKSVLSHRLNLRGKSTRAPIISDAEQVNLBGMVNRKLGRT QFAYLALAEPMFKYVGEKFGPEFLFPGNEBEGDGYILAF VHDEKWKSELIQVNAQNLKLEASIKLPSVPYGFHGTFIHSDLRKQA"

BASE COUNT 627 a 687 c 516 g 568 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,448-148 Length: 2398
 Score: 2172.00 Matches: 428
 Percent Similarity: 78.21% Conservative: 60
 Best Local Similarity: 68.59% Mismatches: 100
 Query Match: 68.95% Indels: 36
 DB: 8 Gaps: 9

US-09-758-269-6 (1-599) x AF190462 (1-2398)

QY 3 SerPheThrAlaThraAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnPro 22
 DB 115 TCAATAACCATGCTTCCACAGCTTCAAATACCTGGATT-----AACACCACACTCCCA 168
 QY 23 ProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSer----- 37
 DB 169 TCTTCTGTCTTCCCTTTAAAGACCTGCTTCCACATTTTCTCTCCACACCTTA 228
 QY 38 LeuProMetAlaSerArgValThrArgLysLeuAsn---ValSerSerAlaLeuHisThr 56
 DB 229 CTTCCTTCAAGAAAGATCTCTTCCACACCAACACCATCATCATCTCTCTCAAAACA 288
 QY 57 ProProAlaLeuHisPheProLysGln-----SerSerAsnSerPro 70
 DB 289 -----CTCACTACCCCAACAGTACCMACCAACATCCATCCACCAACCAAC 339
 QY 71 AlaIleValValProLysAlaLysGlu----- 80
 DB 340 CCAACTCCGATTAAACCCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 399
 QY 81 -----SerAsnThrLysGln-----MetAsnLeuPheGlnArgAlaAla 93
 DB 400 CTTCTCTCCGACCAACCAACCAACCAATTCCTCAAAATGGAATTTCTCGAAGAGCCGCT 459
 QY 94 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeu 113
 DB 460 GCGACAGGCTTGATATGTTGGAGACGGCGCTGGTTTCGCACGAGAGCAACACCGCTC 519

QY 114 ProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGln 133
 DB 520 CCCAAAACGGCGGACCCGAAAGTCCAAATCCCGGGAACCTTCGCGCGGTCGCGAGCAC 579
 QY 134 ProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyr 153
 DB 580 GCGCGGATCAAGCCCTTCGCTGGTGGAAAAATCCCAAAATGCATTCAGCGGCTGTAC 639
 QY 154 ValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePheAspGly 173
 DB 640 GTGCGCAACGGCGGAATCCGCTCTACGAGCCCTGCGCGGCGCACCATCTTCTTCAGCGC 699
 QY 174 AspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPhe 193
 DB 700 GAGCGCATGTCACCGCTGCAAGTTACGAACGGCGGTGCAGCTACGCTGCGCTTC 759
 QY 194 ThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAla 213
 DB 760 ACCGAACACACGCGTCTCGCGAAGAAATCTTTAGGCCCGCCGGTGTTCCTCCGAGGCT 819
 QY 214 IleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaAlaAla 233
 DB 820 ATCGCGGAACCTCACGGCCACTTCGGGCATAGCTCGCTTCTCTTCTACGCTCGCAGC 879
 QY 234 AlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyr 253
 DB 880 CTCCTCCAGCTGTTGATGGGTCCACGGCATGGGGTGGGGAACCGCGGTCTGCTCAT 939
 QY 254 PheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThr 273
 DB 940 TTCAACACCACTCTTGGCCATGTCGGAAGAGATTATACCATACCACTGAGAATACCC 999
 QY 274 ProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThr 293
 DB 1000 TCTAACGGCGACTTAACCACTGTCGCGTTTACGACTTCAACGGCGCACTCAACTCCACA 1059
 QY 294 MetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAsp 313
 DB 1060 ATGATCGCCACCCGAAACTCGACCCCGTCAACGGCGACTTCCACGGCTCAGCTACGAC 1119
 QY 314 ValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerPro 333
 DB 1120 GTCGTTTCAGAGCCTTACCTCAAGTACTTCCGTTTCTCCGCTGACGGGTAAAGTCCGCC 1179
 QY 334 AspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGluAsn 353
 DB 1180 GAGTCGGAATCCCTGAAAGAGCCACCATGATGACGATTCGCAATTAACGGAGAAT 1239
 QY 354 PheValValProAspGlnValValPheLysLeuProGluMetIleArgGlyGly 373
 DB 1240 TTCGTGCTGCTCCCGACCCAGCAGGTGGTCTTCAAACCTAACAGAGATGATCACCGCGGC 1299
 QY 374 SerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAla 393
 DB 1300 TCCCGCGTGTCTACGACAAAGAACAAACCTCACCGTTTGGGATTTTGGATAAGATGCG 1359
 QY 394 GluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrp 413
 DB 1360 AAGGACCGAATCGCATGCGGTGGTTCGACGGCGCGGAGTGTTCGTGTTCCATCTCTGG 1419
 QY 414 AsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThrPro 433
 DB 1420 AACCGTGGGAGGAGCTGAAACCCGAGAGATGTGGTGGTTCCTGCTGATGACCCCT 1479
 QY 434 ProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArg 453
 DB 1480 GCGGACTCTCATTTCAATGATGATGAGATTGAGAGCGGTCTTGTCTGAGATAAGG 1539
 QY 454 LeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGln 473
 DB 1540 CTGAATTCAGGACGGGCAAGTCGATCGCCCGCCCATTTATCTCC---GACCCCAACAA 1596
 QY 474 ValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAla 493

```

Db      1597  GTGAACCTTGAAGCCCGCATGTGTGAACAGAAACAAGCTAGGAGGAAACCTCAGTTCGCT 1656
Qy      494  TyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThr 513
Db      1657  TATCTGGCTCTGGCCGAGCCCTGGCCCAAGTCTCGGGCTTTGCGAAGTTGATTGTTC 1716
Qy      514  ThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPhe 533
Db      1717  AGTGGGGAAGTGAGAAAGTACATGTATGGAGAGAAAGTTCGGTGGGAGGCTCTGTATT 1776
Qy      534  LeuProGlyGluGlyGluGluAspGluGlyTyrIleLeuCysPheValHisAspGlu 553
Db      1777  CTTCC---AACGCGAAGAGAGGTGATGGGTATATCTGGCATCTGTGACAGCG 1833
Qy      554  LysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThr 573
Db      1834  AAAGAGTGAATCCGAGCTGCAGATTGTGAATGCCCAAAATTTAAAGTCGAAGCTTCC 1893
Qy      574  ValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGlyAlaAsp 593
Db      1894  ATCAAACTCCCTCTCGCGTTCCTTATGGTTTTCATGGCACATTTTATTCATTAAAGAT 1953
Qy      594  LeuAlaLysGln 597
Db      1954  TTGAGGAGGCAA 1965

RESULT 12
LOCUS   AX652128 95769 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 1979 from Patent WO03000898.
ACCESSION AX652128
VERSION AX652128.1 GI:29154942
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Chang H.S., Chen W., Cooper B., Glazebrook J., Goff S.A., Hou Y.M.,
Karagiri F., Qian S., Tao Y., Whitham S., Xie Z., Zhu T. and Zou G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 1979 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1. 95769
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 30958 a 16961 c 16682 g 31168 t
ORIGIN

Alignment Scores:
Pred. No.: 3,45e-146 Length: 95769
Score: 2170.50 Matches: 419
Percent Similarity: 80.61% Conservative: 80
Best Local Similarity: 67.69% Mismatches: 87
Query Match: 66.90% Indels: 33
DB: 6 Gaps: 9

US-09-758-269-6 (1-599) x AX652128 (1-95769)
Qy 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeu----- 15
Db 33716 ATGGCTTCTACTACTTACTTCTCTTCTACTTCAACCCCAATTCCTCGACCGAATTTCTCA 33775
Qy 16 --GlyGlyAsnHisThrGlnProLeuProLeuSerSerGlnSerSerAspLeuSerTyr 34
Db 33776 ACTTCATCTTCTCTCTCGAGCCTAAACTTCAATCT-----CTATCTTTT 33920
Qy 35 CysSerSerLeuPro-----MetAlaSerArgValThr 45

```

```

Db      33821  TCCTCAACACTTAGAAACAAGAAACTAGTAGTCCCTGCTATGTTTCTTCTCCGTCAC 33880
Qy      46  ArgLysLeuAsnValSerSerAlaLeuHisThr-----ProProAlaLeuHis 61
Db      33881  AAAAAATCTTCGCTCTGCTTCTCTCCATCTCCGACATTCAAACACACCGTCTGG--- 33937
Qy      62  PheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGluSer 81
Db      33938  -----AAGAAACTATGTAACGATGTCCCAACTTGATC-----CCCAAAACGACGAATCAG 33988
Qy      82  AsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGlu 101
Db      33989  AATCAAAA---TTAAACCCCGGTGCAGAGAACCGCCTATGGTTTATAGACGGGTGGAA 34045
Qy      102  GlyPheLeuValSerHisGlu---LysLeuHisProLeuProLysThrAlaAspProSer 120
Db      34046  AACGCCATGATCTCACAGAGCGCGCGCTCATCCCATCTCTAAACGGCGGATCTCTGCC 34105
Qy      121  ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
Db      34106  GTCCAAATAGCGGGAACCTTCTCCCGTACCGGAGAAACCTGTCTGTCATTAACCTTCCG 34165
Qy      141  ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db      34166  GTGACTGGAACAGTACCAGATGCTTCAAGGAGTTTACGTACAGAAACGGAGCAATCCA 34225
Qy      161  LeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaVal 180
Db      34226  CTTCAAAACAGTCTCCGGCCACCATTTGTTCCAGCGAGACGCTATGGTACACGCTGTC 34285
Qy      181  LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db      34286  CGGTTCGATAACGGCTCGGTAGCTAGCTTGTCCGTTTACCAGAAACAAACCGGTTGGTT 34345
Qy      201  GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db      34346  CAGAAGCTGTAATGGTGTGTCGCTCGCTTCCCAAGCAATTTGGAGAGCTTCGCGACAT 34405
Qy      221  ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
Db      34406  TTAGAAATCGCAAGCTTATGCTCTTCAATACCGCGCGGCTATTGGGTGTAGTCGACCCG 34465
Qy      241  AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
Db      34466  ACCGAGGAGCTCGGTGTCGCTAACCGCGTTAGTTATTATTTCAATGTCTCTTAGCC 34525
Qy      261  MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
Db      34526  ATGTCGGAAGACGATTTACCGTACCATGTCAAAGTTACTCAAACCGGAGATTTAGAAACT 34585
Qy      281  ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
Db      34586  TCGGGTCGGTACGATTTTCAGCGTCAGTTAAATCAACATGATAGCCACCCCGAANAATC 34645
Qy      301  AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysPheProTyrLeu 320
Db      34646  GATCCCGAACCAGAGAACTCTTCGCTTTAAGCTACGACGCTGCTTTCAAAGCTTACTTA 34705
Qy      321  LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
Db      34706  AAATATCTTCAATTCATCCATCCGACGTCAGAAATCACCAGACGTCGAGATTCGCGTTGAT 34765
Qy      341  GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
Db      34766  CAACCAACGATGATCCAGATTTCCGATCACTGAGATTTTCGTAGTATTCACAGACCAA 34825
Qy      361  GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValValTyrAspLys 380
Db      34826  CAAGTGGTATTCGATTCGCGAGATGATAGAGGTGTTCTCCGCTGGTTTACACGAG 34885
Qy      381  AsnIleValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db      34886  AAGAAGAAATCAAGATTCGGGATTTTGAATAAAACCGGAAAGATGCTTCGTCGATTCAA 34945

```

QY 401 TTPileaspAlapProaspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu 420
 Db 34946 TGGATCGAAGTACACAGTTGTTTGTTCATCTATGGACTCTTGGAGACACCGAA 35005
 QY 421 ThrAspGluValValValledGlySerCysMetThrProProAspSerIlePheAsnGlu 440
 Db 35006 ACAGACGAGGTGTCTGTGATTGATCATGATGACGACCGACCTGATTCAATTTTCACGAA 35065
 QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 Db 35066 CACGACGAAACACTTCAGAGTGTGTTCGAGAGTAAGCTTAACCTTAACACGAGGAA 35125
 QY 461 SerThrArgArgProIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
 Db 35126 TCAACACGAGTACCGGTATCTTCG-----GAACAAGTTAATCTCGAAGCTGGATG 35176
 QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyLeuAlaLeuAlaGluPro 500
 Db 35177 GTCAACCGGAATTTATTAGTAGAATAACCCGGTATGCTTATCTGCTTTAACCGAACCG 35236
 QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 35237 TGGCCTAAGTGTCCGGTTCGCTAAGTGGACTTATCGACCGAGAGATTGCAAGTAT 35296
 QY 521 LeuTyArgAspAsnArgTyArgGlyGluProLeuPheLeuProGlyGlyGlyGlu 540
 Db 35297 ATTACCGGAGGAGGAATAACGAGGAGAGCCCTCTGTTCTACCTTCGGTGACCGAGAA 35356
 QY 541 GluAspGluClyTyTrileuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
 Db 35357 GAAGACGAGGTACATAATGTTGTTTCACGAGGAGGAAGTGAATTCGGAATC 35416
 QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 35417 CAACCTATAACCGCGTTAATAGAGCTTGAAGCTACGTAACGCTTCGCTCGAGAGTG 35476
 QY 581 ProTyGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
 Db 35477 CTTAGCTTTCACGGAAGCTTATCAGTAGGAAGATTTATCGAAGCAAGCTTTG 35533
 RESULT 13
 AC013430
 LOCUS
 DEFINITION 95769 bp DNA linear PLN 26-MAY-2000
 Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
 I, complete sequence.
 AC013430
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 95769)
 REFERENCE
 AUTHORS
 Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
 Shinn,P., Altafi,H., Bei,Q., Chin,C., Choi,J., Conn,L., Conn,L.,
 Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
 Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N.,
 Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
 Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A.,
 Theologis,A., and Ecker,J.R.
 TITLE
 Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
 I
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 95769)
 Ecker,J.R.
 Direct Submission
 TITLE
 Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th Street and
 Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 JOURNAL
 REFERENCE
 3 (bases 1 to 95769)
 Ecker,J.R.

Direct Submission
 Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th Street and
 Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 4 (bases 1 to 95769)
 REFERENCE
 AUTHORS
 Chao,Q., Brooks,S., Buehler,E., Chao,Q.,
 Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C.,
 Choi,J., Conn,L., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
 Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
 Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
 Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
 Yu,G., Davis,R., Federspiel,N., Theologis,A., and Ecker,J.R.
 Direct Submission
 Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th and
 Hamilton Walk, Philadelphia, PA 19104-6018, USA
 5 (bases 1 to 95769)
 REFERENCE
 AUTHORS
 Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C.,
 Choi,J., Choi,E., Conn,L., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
 Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
 Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
 Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
 Yu,G., Davis,R., Federspiel,N., Theologis,A., and Ecker,J.R.
 Direct Submission
 Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th and
 Hamilton Walk, Philadelphia, PA 19104-6018, USA
 On May 26, 2000 this sequence version replaced gi:6921155.
 COMMENT
 FEATURES
 source
 1..95769
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="F3F9"
 join(71..402,574..649,733..933)
 /note="similar to oxidosqualene cyclase dbj|BAA33462.1"
 /codon_start=1
 /evidence=not_experimental
 /product="F3F9.1"
 /protein_id="AAF71792.1"
 /db_xref="GI:8052528"
 /translation="MGHVOLLTFCYFPPFFSFLFLFLNKNVILISVDITVGT
 FELAKVEQTPPVKVEDGSGTHETMEYGEFLSSRLCRPPMTVGLSKSLDSS
 PLHWYSSFGHLEIEFDAREHEILRYIYRHLNDGGGLHVECKSMFCTALNYI
 CLRIRSGFDGREGYACKLISKMPWIVTLFYNNESCSPSI"
 complement(join(3438..3746,3868..4032,4110..4664))
 /note="similar to F-box domain gb|AAF14684.1"
 /codon_start=1
 /evidence=not_experimental
 /product="F3F9.2"
 /protein_id="AAF71813.1"
 /db_xref="GI:8052549"
 /translation="MLASWICSLTSMKTKNLIANPNRFAFTGDNVLKSTFTFN
 HPSLRTLSYQILQDAQANNCKNLYLDLFOIFAEVDVFNVEVSSCSLEVLILQI
 IFFNPSGHLKDHKTLILSNQDSIEVRAARLDLSIEYSPESDNVLEIPRL
 OYFNPSVAGRLPHTSINTSCPPQESNGMVKDNTYATSPASLSLDKNPREV
 EVLKILAVTDEKIEVEISFNKNPLGECNSDGAQNNLMEKLEPPFNADPRVDT
 IWMNFKGSNKEQFALASRFVMQKTMVKMKWIKTSPEDEKKKEIAAVALKELPKGN
 EELSIECF"
 join(7138..7400,7491..7887)
 /note="unknown protein; similar to EST gb|AI999747.1"
 /codon_start=1
 /evidence=not_experimental
 /product="F3F9.3"
 /protein_id="AAF71793.1"
 /db_xref="GI:8052529"
 /translation="METGLNLIKLSCLSVVGVSAQAAPWNPSPGPRGTCDHVE
 CPTVKLYEAGYGFIRYDAALWTSPIPSLSWTOATKTGFERLYLEGONKSNVK
 KMYTAPVIAQTGPRSVYTSVLIPKKNQONPCADDLHVRSKTYPTVAVRQICGVYS
 NNVADEAAALMESLRSNWLIPKKNLPPAYFLAVYNPPSTHTTARVINELWPFN
 M"

```

CDS
join(8524..8759,8846..9287)
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.4"
/protein_id="AAF71794.1"
/db_xref="GI:8052530"
/translocation="MEALISIFKLSLLTLLSGGSLDLSLPPNCNRAECPSYEVVHA
GNGYEIRHYNTTWIETPIQDLSNEAGSNQWLSIDYMGNNPNYHORIEIALPYIT
OVSONSLSTFTVSPFVKAPQDPDPPGNNLHVORWDSYVAVKQISYVADHKGQOVA
ELKASLQGTWAAIKESKRETTGGGSAWAYTVAQFSWPFQMSQRVNEIWFPEMDEE
TVSIVIQ"
CDS
join(13444..13814,13956..14310,14399..14662)
/notes="similar to dioxygenase protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.5"
/protein_id="AAF71795.1"
/db_xref="GI:8052531"
/translocation="MAVLKAPVAPKSGSLPLVIDMSPEKHALVKACEFGPFKV
INHGSAELVLEHETVDFSLPSEKTOVAGYFGYSGKISGRNGDVGWVEYLLMN
ANDSGGFLPPLSKSPGTRNALREYTSVRKMTFDVLEKIDTGLKPKRNTLSKL
VQSDNTDSIURLNHYPCPLSNKKTNGKNVIGFHEHTDPQIIIVLSRNTSGLTQNL
NDGWSLVPDPHSPFFNVGDSLQVMTNGRFSVRHVLANKCKSRVSIYFAGPSLT
ORLAPTCLIDNEDELYREFTWSEYKNTSYNSRLSDNRLQOQFERKTIKNLLN"
complement(join(15755..16715,16810..16829))
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.6"
/protein_id="AAF71812.1"
/db_xref="GI:8052548"
/translocation="MKPRLGIRGSELPOQSRPLATSLTSSDPHLSRPTDRSP
KUGDRRSGGPHDPLSQKLGSRISGLESQLOCAQAEELRLKQALAKAAKRR
AQELHRKSGPHDPLSQKLGSRISGLESQLOCAQAEELRLKQALAKAAKRR
INVLKARYLEKERSVLEENETLKQDKKTSTEMSCAKAKEDATASVSGIGSELE
ESNETAKLYLKEVEEAETLEAEMKLLKVQTEQWKAADAADAAALSLGVEMGRF
SEOGSMKEHFPACRFVSGPMADDSDGSGKSSGKMGFDLWKKGQK"
complement(join(19162..19689,19767..19833,20044..20164,
20234..20378,20482..20623,20693..20738,20826..20877,
20965..21156))
/notes="unknown protein; similar to ESTs gb|T41672.1,
gb|A1992710.1, and gb|AA586078.1"
/codon_start=1
/evidence=not experimental
/product="P3F9.7"
/protein_id="AAF71811.1"
/db_xref="GI:8052547"
/translocation="MGNLGRKQVVEERYTKQGLYVNVKDVVKLRLVIESKLAP
CYGDDSDHDLSECFICFLYPLNRSRCMKSICTEFLQMNPNARPQKFLIN
LMGFPTVSLTMRCPFCPTPNAYEYRGVSKSEKIEQVEEIVIAKIMRQKE
MODDEKQKRLSCSSTGAMTEYSGTGLFVFPFAAISYNLSMDGGEIAPSQ
NASVVRQHSRPRGNEVDVLELMVMEAIWLSVQETGTQNSASGETISRQVYT
DNHSYLSRPRVPIVEPATPSSSGSLSCAISALAEQVMVGESSSHNNHNVVSSY
SMUPGNCDSYDIEQVDGIDNHHHHHHHMETGSSNSYVSYWTGSGFHNFPFPP
PLVIVPSEBQMMMAVAMAEVHATTTCAPTEVTWQ"
complement(join(22808..23063,23118..23244,23511..23706))
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.8"
/protein_id="AAF71810.1"
/db_xref="GI:8052546"
/translocation="MSGRGKVPKPMKVFINTQYVETDARSFKTVVQELTGNKATVA
AGPDPSPSADGRCYDGSGKIGPITNVVFWKVMYSLVSPNPNFNRYSFPHLFPK
RTMKPTNYQNKKECVTKALKATSLRSQASTGFSLWGVATQVLQGLTFPSIFPI
YLVFMFFSCIKMFLTKRLISLKLIVSLTV"
join(24752..26907,27020..27871,27923..28153)
/notes="similar to exopolysaccharonase precursor"
/codon_start=1
/evidence=not experimental
/product="P3F9.9"
CDS
join(8524..8759,8846..9287)
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.4"
/protein_id="AAF71794.1"
/db_xref="GI:8052530"
/translocation="MEALISIFKLSLLTLLSGGSLDLSLPPNCNRAECPSYEVVHA
GNGYEIRHYNTTWIETPIQDLSNEAGSNQWLSIDYMGNNPNYHORIEIALPYIT
OVSONSLSTFTVSPFVKAPQDPDPPGNNLHVORWDSYVAVKQISYVADHKGQOVA
ELKASLQGTWAAIKESKRETTGGGSAWAYTVAQFSWPFQMSQRVNEIWFPEMDEE
TVSIVIQ"
CDS
join(13444..13814,13956..14310,14399..14662)
/notes="similar to dioxygenase protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.5"
/protein_id="AAF71795.1"
/db_xref="GI:8052531"
/translocation="MAVLKAPVAPKSGSLPLVIDMSPEKHALVKACEFGPFKV
INHGSAELVLEHETVDFSLPSEKTOVAGYFGYSGKISGRNGDVGWVEYLLMN
ANDSGGFLPPLSKSPGTRNALREYTSVRKMTFDVLEKIDTGLKPKRNTLSKL
VQSDNTDSIURLNHYPCPLSNKKTNGKNVIGFHEHTDPQIIIVLSRNTSGLTQNL
NDGWSLVPDPHSPFFNVGDSLQVMTNGRFSVRHVLANKCKSRVSIYFAGPSLT
ORLAPTCLIDNEDELYREFTWSEYKNTSYNSRLSDNRLQOQFERKTIKNLLN"
complement(join(15755..16715,16810..16829))
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.6"
/protein_id="AAF71812.1"
/db_xref="GI:8052548"
/translocation="MKPRLGIRGSELPOQSRPLATSLTSSDPHLSRPTDRSP
KUGDRRSGGPHDPLSQKLGSRISGLESQLOCAQAEELRLKQALAKAAKRR
AQELHRKSGPHDPLSQKLGSRISGLESQLOCAQAEELRLKQALAKAAKRR
INVLKARYLEKERSVLEENETLKQDKKTSTEMSCAKAKEDATASVSGIGSELE
ESNETAKLYLKEVEEAETLEAEMKLLKVQTEQWKAADAADAAALSLGVEMGRF
SEOGSMKEHFPACRFVSGPMADDSDGSGKSSGKMGFDLWKKGQK"
complement(join(19162..19689,19767..19833,20044..20164,
20234..20378,20482..20623,20693..20738,20826..20877,
20965..21156))
/notes="unknown protein; similar to ESTs gb|T41672.1,
gb|A1992710.1, and gb|AA586078.1"
/codon_start=1
/evidence=not experimental
/product="P3F9.7"
/protein_id="AAF71811.1"
/db_xref="GI:8052547"
/translocation="MGNLGRKQVVEERYTKQGLYVNVKDVVKLRLVIESKLAP
CYGDDSDHDLSECFICFLYPLNRSRCMKSICTEFLQMNPNARPQKFLIN
LMGFPTVSLTMRCPFCPTPNAYEYRGVSKSEKIEQVEEIVIAKIMRQKE
MODDEKQKRLSCSSTGAMTEYSGTGLFVFPFAAISYNLSMDGGEIAPSQ
NASVVRQHSRPRGNEVDVLELMVMEAIWLSVQETGTQNSASGETISRQVYT
DNHSYLSRPRVPIVEPATPSSSGSLSCAISALAEQVMVGESSSHNNHNVVSSY
SMUPGNCDSYDIEQVDGIDNHHHHHHHMETGSSNSYVSYWTGSGFHNFPFPP
PLVIVPSEBQMMMAVAMAEVHATTTCAPTEVTWQ"
complement(join(22808..23063,23118..23244,23511..23706))
/notes="hypotheetical protein"
/codon_start=1
/evidence=not experimental
/product="P3F9.8"
/protein_id="AAF71810.1"
/db_xref="GI:8052546"
/translocation="MSGRGKVPKPMKVFINTQYVETDARSFKTVVQELTGNKATVA
AGPDPSPSADGRCYDGSGKIGPITNVVFWKVMYSLVSPNPNFNRYSFPHLFPK
RTMKPTNYQNKKECVTKALKATSLRSQASTGFSLWGVATQVLQGLTFPSIFPI
YLVFMFFSCIKMFLTKRLISLKLIVSLTV"
join(24752..26907,27020..27871,27923..28153)
/notes="similar to exopolysaccharonase precursor"
/codon_start=1
/evidence=not experimental
/product="P3F9.9"

```

Db 34226 CTTCAAAACAGCTCTCCGCCACCATTTGTTTCGACGGAGACGGTATGTTACACGCTGTC 34285
 Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
 Db 34286 CGGTTCAATACAGCTCGGTAGTACGCTTGTGCGTTTACCGAAACAAACCGGTTGGTT 34345
 Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
 Db 34346 CAAGACGTGAATGTTGTTGCGGTTTCCCAAGCAATTCGAGAGCTTCACGGACAT 34405
 Qy 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaIleGlyIleValAspPro 240
 Db 34406 TTAGGAATCGCAAGCTTATGCTCTTCAATACCGCGGCTATTGCGGTAGTTCGACCGG 34465
 Qy 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
 Db 34466 ACCGAGAGCTCGGTGTCGTAAACCGCGGTTTATTTTCAATGTCATCTCTTAGCC 34525
 Qy 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
 Db 34526 ATGTCGGAAGACGATTTACCGTACCATGTCAAAGTTACTCAACCGGAGATTTAGAACT 34585
 Qy 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
 Db 34586 TCGGTCGGTACGATTTCCGACGTCAGTTAAATCAACATGATGAGCCACCCGAAATC 34645
 Qy 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
 Db 34646 GATCCCGAAACCCGAGAACTCTTCGTTTAAAGTACGACGTCGTTTCAAAAGCCTTACTTA 34705
 Qy 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
 Db 34706 AAATACTTCAGATTCACATCCGACGGTGAGAAATCACGACGTCGAGATTCGCTTGAT 34765
 Qy 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360
 Db 34766 CAACCAACGATGATCCAGATTCGCGATCACTGAGATTTTCGTAGTATTCAGACCAA 34825
 Qy 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
 Db 34826 CAAGTGGTATTCGATTCGCGAGATGATAAGAGTGCTTCTCCGGTGGTTTACGACGAG 34885
 Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
 Db 34886 AAGAAGAATCAAGATTCGGATTTTGATTAACCAACCGGAAAGATGCTTCGTCGATCAA 34945
 Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTyrPheAsnAlaTyrGluProGlu 420
 Db 34946 TGGATCGAAGTACCAAGATTTGTTTGTTCATCTATGAACTCTTGGGAAGAACCGAA 35005
 Qy 421 ThrAspGluValValIleGlySerCysMetThrProAspSerIlePheAsnGlu 440
 Db 35006 ACAGACGAGTTGTGTTGATTCGATTCATGATGATGACGCCACCTGATTCATTTTCAACGA 35065
 Qy 441 SerAspGluAsnLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 Db 35066 CACGACGAACACTTCAGAGTCTTTGTCGAGATAAGCTAAACCTTAAACACGAGGAA 35125
 Qy 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
 Db 35126 TCACACGCTAGACCGGTTATCTCG-----GACCAAGTTAACTCGAAGCTGATG 35176
 Qy 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
 Db 35177 GTCAACCGGAATTTATTAGGTAGAAAAACCGGTATGCTTATCTGGCTTAAACCGAACCG 35236
 Qy 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 35237 TGGCCTAAAGTGTCCGGTTTCGCTAAAGTGGACTTATCCACCGAGAGATTCGAAGAT 35296
 Qy 521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGluGlyGlu 540
 Db 35297 ATTTCGAGAGAGGGGAATACGAGGAGAGCGCTCTGTTTCTACCTTCGGTACGCGAGAA 35356

Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTyrLysSerGluLeu 560
 Db 35357 GAAGACGGAGTTACATAATGGTCTTCTGTTACGACGAGGAGAGGTGAAATCGAACTC 35416
 Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 35417 CAATCATAAACGCGCTTAATATAGCTTGAAGCTACCGTAAACGCTTCGCTCGAGAGTG 35476
 Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
 Db 35477 CCTTACGGTCTCCAGGAACTTCATCAGTAAGAAAGATTTATCGAAGCAAGCTTTG 35533
 RESULT 14
 AXI48312
 LOCUS AXI48312 1839 bp DNA linear PAT 08-JUN-2001
 DEFINITION Sequence 11 from Patent EP1116794.
 ACCESSION AXI48312
 VERSION AXI48312.1 GI:14347199
 KEYWORDS Vigna unguiculata (cowpea)
 SOURCE Vigna unguiculata
 ORGANISM Vigna unguiculata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.
 REFERENCE 1
 AUTHORS Iuchi, S., Kobayashi, M. and Shinozaki, K.
 TITLE Transgenic plants carrying neoxanthin cleavage enzyme gene
 JOURNAL Patent: EP 1116794-A 11 18-JUL-2001;
 Riken (JP)
 FEATURES
 source
 1..1839
 /organism="Vigna unguiculata"
 /mol_type="genomic DNA"
 /db_xref="taxon:3917"
 CDS
 1..1839
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC41197.1"
 /db_xref="GI:14347200"
 /translation="MPSASNTWFNATLSPPEKDLSTSSPTNLLPLRKTSSNTTIT
 CSQTLHPFKQPTSTSTATTTTTPFKTTTTPPRETNFSDTNPLOPKW
 NFKQAATAADLVETALVSHRKHPLADPRVQIAGNFAPVPEHAADGGLFVWGK
 IPKIDGVVANGANPLYPVAGHHFDDGMVHAVKFTNGAASACRFTQRLSQE
 KLGPRPFPKAIHELHSGIARLLFYARGLFGVDSQGMVANGLVFNHLLA
 MSEDLLPVHRTTPNGDLTVGRYDENGQNSTIAHFKLPDVGDLHALSYDIQKP
 VLKYPSPDGKSPDVEIPLEKPTMHDFALTENFVVVDDQVVFVKTENIGTSPV
 YDKNKTBRFGILHKNAKDANMKWIDAPDCFCFHLWNAWEEPETEEVVTGSKTQPA
 DSIFNECESLKSLSLRDLNLRKSTRRIISDAEQVNLKAGVWENKLGKRTQPA
 YLALAPWPKVSGFAKVDLLSGEVKKYMGEEKFGEFLFNGKEDDGYILAFVHD
 EKEWSELIQVNAONLKLEASIKLPSRPVYFGHTFIHSKDLRQQA"
 BASE COUNT 447 a 595 c 442 g 355 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,78e-148 Length: 1839
 Score: 2168.50 Matches: 423
 Percent Similarity: 78.44% Conservative: 61
 Best Local Similarity: 68.56% Mismatches: 98
 Query Match: 68.84% Indels: 35
 DB: 6 Gaps: 8
 US-09-758-269-6 (1-599) x AXI48312 (1-1839)
 Qy 8 AlaAlaValSerGlyArgTyrLeuGlyGlyAsnHisThrGlnProProLeuSerSerSer 27
 Db 7 TCATCAGCTTCAACACACTTGGTTTAAACGCCACACTCCCATCTCCCTCCCTTCAAGACCTA 66
 Qy 28 GlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLys 47
 Db 67 CCTTCCACATCT---TCTCCACACAACTTACTTCTCTTA-----AGGAAACATCTCTCT 117

QY 48 LeuAsn---ValSerSerAlaLeuHisThrProAlaLeuHisPheProLysGln--- 65
 Db 118 TCCAAACACCATCATGTTCTCCCTCAACA-----CTCCACTTCCCAACAGTAC 168
 QY 66 -----SerSerAsnSerProAlaIleValValLysProLysAlaLysGluSer 81
 Db 169 CAACCAACATCCACATCCATCCACAGCCACACACACACACACCCCAATCAAAACT 228
 QY 82 AsnThr----- 83
 Db 229 ACCACCATCACACACACACCGCCAGGAAACCAACCCCTCTCTGACACCAACCAA 288
 QY 84 -----LysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
 Db 289 CCATTACCTCAAAATGGAACTTTCTCCAGAAAGCGCTGCCACCGCCTTGACCTGTC 348
 QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
 Db 349 GAAACGGCGCTGCTTCGACGAGCGCAACACCCCGCTCCCAACAAACGGCGACCCGAGG 408
 QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
 Db 409 GTCCAAATCGCCGGGAATTCGCGCGGTGCGGAGCATGCCGCGATCAAGGACTCCCG 468
 QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTy-ValArgAsnGlyAlaAsnPro 160
 Db 469 GTGGTCGGAATAATCCCAATCCATGACGGGTGACGTGCGGCACCGCTCCCAATCCG 528
 QY 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
 Db 529 CTCCTACGAGCCTGTGCGCGGACCACTTCTTCGACGCGAGCGGCATGCTCCACGCGGTG 588
 QY 181 LysPheGluHisGlySerAlaSerTyAlaCysArgPheThrClnThrAsnArgPheVal 200
 Db 589 AAGTTCAGACGCGCGCGGACCTACGCTCGCGCTTCACGAGACGAGCGTCTCTCG 648
 QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyLeuHisGlyHis 220
 Db 649 CAGGAGAAATCTCTAGCGCGCGGTGTTCCCGAAGGCCATCGGGAGCTCCACGCGCCAC 708
 QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyAlaArgAlaAlaAlaGlyIleValAspPro 240
 Db 709 TCCGGCATCGCGGGCTCTCTCTCTACGCGCGGTCTCTCTCGGGCTCGTGTAGTGG 768
 QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyPheAsnGlyArgLeuLeuAla 260
 Db 769 TCCAGGCGCATGGCGGTGGGACCGCGGTCTCTCTACTTCAACACCACTCTTGGCC 828
 QY 261 MetSerGluAspAspLeuProTyGlnValGlnIleThrProAsnGlyAspLeuTyThr 280
 Db 829 ATGTCCGAAGACGATTTACCTTACCGTACGAGATCACCCCTAACGCGGACTTAAACACC 888
 QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
 Db 889 GTTGGCGGTACGACTTCAAGCGGCGAGTCACTCAACATGATCGCCCAACCGAACTG 948
 QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyAspValValSerLysProTyLeu 320
 Db 949 GACCCCGTCAGCGCGGACCTCCAGCGCTCAGGTACGATGATTCAGAGCGCTTACCTC 1008
 QY 321 LysTyThrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
 Db 1009 AAGTACTTCGTTTCTCCCCCGCGCGGTCAAGTCCCCCGAGCTGGAAATCCCCCTGAAG 1068
 QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360
 Db 1069 GAGCCCAACCATGATGCACGATTTCCGCAATACGAGAAATTCGTGCTGCTCCCGGACCCAG 1128
 QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyAspLys 380
 Db 1129 CAGGTGGTCTTCAAACTAACCGAGATGATCACCGCGGGTCCCCCGGTGCTACGACAG 1188
 QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyAlaGluAspSerSerAsnIleLys 400

Db 1189 AACAAACCTCAGCGTTGGGATTTCGCACAGAATGCGAAGACCGCAATGCGATGCGG 1248
 QY 401 TTPleAspAlaProAspCysPheHisLeuTyPheAsnAlaIleTyPheGluProGlu 420
 Db 1249 TGAATCGACGCGCGGATTTCTCTCCACTCTCGAAGCGGTGGGAGGAGCCGAA 1308
 QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
 Db 1309 ACCGAGGAGTGTGGTGTGATTGGTCTCTGCATGACCCCTGCGGACTCCATTTTCAACGAA 1368
 QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
 Db 1369 TGGAGGAGATTTGAAGACGCTGTGACAGTAAGCTGAACCTTAGACCGCGCAAG 1428
 QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480
 Db 1429 TCCACTCGCGCGCCCTATATCTCC---GACGCGCAACAGTGAACCTGGAAGCGCGCATG 1485
 QY 481 ValAsnArgMetLeuGlyArgLysThrLysPheAlaTyPheAlaLeuAlaGluPro 500
 Db 1486 GTGAACAGAAACAGCTCGGAAGAGAGACCCAGTTGCGTATCTGGCTCTGGCGGAGCCC 1545
 QY 501 TTPProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
 Db 1546 TGGCCCCAAGTCTCGGCTTTTGGAAAATGATTGCTGAGTGGGGAAGTGAAGAGATAC 1605
 QY 521 LeuTyGlyAspAsnArgTyGlyGluProLeuPheLeuProGlyGluGlyGlu 540
 Db 1606 ATGTATGAGAAGAAAGTTCGGTGGGAGCCCTCTGTCTTCTCC---AACGCGCAAAA 1662
 QY 541 GluAspGluGlyTyIleLeuCysPheValHisAspGluLysThrTyPheSerGluLeu 560
 Db 1663 GAACGCGTGGTATATTCGGCATTCGTCACGACGAGAAAGTAATGAAATCCGAGCTG 1722
 QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
 Db 1723 CAGATTGTGAATGCCAAAATTTAAAGCTCGAAGCTTCCATCAAACTCCCTCTCGTGT 1782
 QY 581 ProTyGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597
 Db 1783 CCTACGGTTTTCATGAACTTTCATTCATTCAGAGGATTTGAGGAACAA 1833
 RESULT 15
 BD017434 1839 bp DNA linear PAT 27-AUG-2002
 LOCUS Transgenic plant using neozanthine cleaving enzyme gene.
 DEFINITION BD017434
 ACCESSION BD017434
 VERSION BD017434.1 GI:22558610
 KEYWORDS JP 2001258579-A/6. (cowpea)
 SOURCE Vigna unguiculata
 ORGANISM Vigna unguiculata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.
 REFERENCE 1 (bases 1 to 1839)
 AUTHORS Iuchi,K., Kobayashi,M. and Shinozaki,K.
 TITLE Transgenic plant using neozanthine cleaving enzyme gene
 JOURNAL Patent: JP 2001258579-A 6 25-SEP-2001;
 THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
 COMMENT OS Vigna unguiculata
 PN JP 2001258579-A/6
 PD 25-SEP-2001
 PF 11-JAN-2001 JP 2001003476
 PI KIYOSHI IUCHI,MASATOMO KOBAYASHI,KAZUO SHINOZAKI PC
 C12N15/09,A01H5/00,C12N5/10,C12N5/00,C12N5/00 CC
 Transgenic plant using neozanthine cleaving enzyme gene FH key
 FT CDS Location/Qualifiers
 1..1839 Location/Qualifiers
 /organism="Vigna unguiculata"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:17:43 ; Search time 36 Seconds
(without alignments)

3037.587 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGNHT.....VPYGFHGTFIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	US-09-758-269-6
2	2280.5	72.4	605	10	Sequence 6, Appl
3	2168.5	68.8	612	10	Sequence 16, Appl
4	1991	63.2	583	10	Sequence 12, Appl
5	1930	61.3	604	10	Sequence 2, Appl
6	1663.5	52.8	577	10	Sequence 14, Appl
7	968	30.7	595	10	Sequence 10, Appl
8	938	29.8	538	10	Sequence 4, Appl
9	937	29.7	538	10	Sequence 8, Appl
10	936	29.7	538	10	Sequence 18, Appl
11	411.5	13.1	505	8	Sequence 33, Appl
12	379	8.9	177	12	Sequence 22, Appl
13	260	8.3	526	15	Sequence 66, Appl
14	247	7.8	506	15	Sequence 1, Appl
15	238.5	7.6	549	12	Sequence 4, Appl
					Sequence 19, Appl

16	232.5	7.4	532	12	US-10-168-517-17	Sequence 17, Appl
17	207.5	6.6	556	12	US-10-168-517-21	Sequence 21, Appl
18	185.5	5.9	529	15	US-10-053-192-5	Sequence 5, Appl
19	150	4.8	620	12	US-10-168-517-2	Sequence 2, Appl
20	107	3.4	26926	10	US-09-759-508B-2	Sequence 83, Appl
21	106.5	3.4	2701	15	US-10-171-311-83	Sequence 77, Appl
22	106	3.4	412	10	US-09-922-261-77	Sequence 784, Ap
23	105.5	3.3	1381	12	US-10-032-585-7784	Sequence 36, Appl
24	105	3.3	1118	12	US-10-410-681-36	Sequence 37319, A
25	105	3.3	5701	9	US-09-864-761-37319	Sequence 32438, A
26	105	3.3	5701	12	US-10-029-386-32438	Sequence 4, Appl
27	104.5	3.3	623	12	US-10-203-351-4	Sequence 6, Appl
28	103.5	3.3	1311	15	US-10-103-377C-6	Sequence 13187, A
29	103	3.3	1156	9	US-09-815-242-13187	Sequence 135, App
30	101.5	3.2	1886	12	US-10-315-515-135	Sequence 5107, Ap
31	101.5	3.2	4342	9	US-09-815-242-5107	Sequence 2, Appl
32	99.5	3.2	383	12	US-10-251-503-2	Sequence 1119, Ap
33	99.5	3.2	2732	12	US-10-238-075-1119	Sequence 252, App
34	99.5	3.2	2834	12	US-10-085-959-252	Sequence 428, App
35	98.5	3.1	1277	11	US-09-934-455-428	Sequence 21, Appl
36	98.5	3.1	2828	10	US-09-905-129-21	Sequence 43, Appl
37	98.5	3.1	2828	12	US-10-301-822-49	Sequence 126, App
38	98.5	3.1	2828	12	US-10-032-189-126	Sequence 54, Appl
39	98.5	3.1	2828	15	US-10-176-847-54	Sequence 110, App
40	98.5	3.1	2828	15	US-10-177-293-110	Sequence 2, Appl
41	98.5	3.1	748	11	US-09-272-975-2	Sequence 58, Appl
42	98	3.1	753	11	US-09-272-975-58	Sequence 3692, Ap
43	98	3.1	411	10	US-09-738-626-3692	Sequence 40, Appl
44	97.5	3.1	1031	12	US-10-410-681-40	
45	97.5	3.1				

ALIGNMENTS

RESULT 1

US-09-758-269-6
; Sequence 6, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-758-269-6

Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 7.9e-309;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASFTATAVSGRWLGNGNHTOPSSQSSDLSYCSSLPMSRVTRKLVSSALHTPPAL	60
DB	1	MASFTATAVSGRWLGNGNHTQPPLSSQSSDLSYCSSLPMSRVTRKLVSSALHTPPAL	60
QY	61	HPFKQSSNPAIVVWPKAKESNTKQNLFORAAAAALDAAGFLVSHKHLPLKPTADPS	120
DB	61	HPFKQSSNPAIVVWPKAKESNTKQNLFORAAAAALDAAGFLVSHKHLPLKPTADPS	120
QY	121	VQIAGNAPVNEQVPRNLVVVGKLPDSIKGVYVNGANPLPEPVTGHHFFDGDGWHAV	180

Db 121 VQIAGNAPVNEQVRNL PVGKLPDSIKGVYVRNGANPLHBPVTVGHFFDGDGMVHAV 180
Qy 181 KFEHGSASYACRFTQTNRVQEROLGRPVFPKPAIGELHGHGTGIALMLFYARAAAAGIVDP 240
Db 181 KFEHGSASYACRFTQTNRVQEROLGRPVFPKPAIGELHGHGTGIALMLFYARAAAAGIVDP 240
Qy 241 AHGTGVANAGLVYFNGRLAMSDDLQVQITPVGDLKTVGRFDFDGOLESIMIAHPKV 300
Db 241 AHGTGVANAGLVYFNGRLAMSDDLQVQITPVGDLKTVGRFDFDGOLESIMIAHPKV 300
Qy 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
Db 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
Qy 361 QVVFKLPEMIRGGSPVYDKNKVARFGLDKYADSSNIKWIDAPDCCFCHLNNAAEPE 420
Db 361 QVVFKLPEMIRGGSPVYDKNKVARFGLDKYADSSNIKWIDAPDCCFCHLNNAAEPE 420
Qy 421 TDEVVIGSCMTPTDPSIFNESDENLKSILRLNLKTGSTRRPIISNEDQOVNLEAGM 480
Db 421 TDEVVIGSCMTPTDPSIFNESDENLKSILRLNLKTGSTRRPIISNEDQOVNLEAGM 480
Qy 481 VNRNMLGKTKFAYLALAEAPKPVSGFAKVDLTGTVKHLVGNRYGGSPFLPQEGGE 540
Db 481 VNRNMLGKTKFAYLALAEAPKPVSGFAKVDLTGTVKHLVGNRYGGSPFLPQEGGE 540
Qy 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYGFHGTGADDLAKQV 599
Db 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYGFHGTGADDLAKQV 599

RESULT 2

US-09-758-269-16

; Sequence 16, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; TITLE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 16

; LENGTH: 605

; TYPE: PRT

; ORGANISM: Lycopersicon esculentum

US-09-758-269-16

Query Match 72.4%; Score 2280.5; DB 10; Length 605;
Best Local Similarity 71.7%; Pred. No. 5.6e-221;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

Qy 5 TATAVSGWLGNGHTQPPPLSSQSSDLSYCS---SLPWASVTRKLVNVSALHTPPALH 61
Db 3 TTTSHATNTWI-----KTKLSMPSSKEFGFASNSISLKNQHNROSLNINSLOAPPLH 57
Qy 62 FPKQSN--SPA--IVVKEKAKESN-----TKQMLLQFAAAAALDAAEGEIVSHEKLP 112
Db 58 FPKQSNYQTPKNNIISHPKQENNNSSSSSTSKNLVQKAAAWALDAVESALTKELEHP 117
Qy 113 LPKTADPSVQIAGNFAVPNEQVRNLPVVGKLPDSIKGVYVRNGANPLHBPVTVGHFFD 172
Db 118 LPKTADPRVQISGNFAPVNPENFVQSLPVTGIPKVCQGVYVRNGANPLFEPETAGHFFD 177

Qy 173 QDGMVHAVKFEHGSASYACRFTQTNRVQEROLGRPVFPKPAIGELHGHGTGIALMLFYAR 232
Db 178 QDGMVHAVQFNGSASYACRFTETERLVQEKALGRPVFPKPAIGELHGHGTGIALMLFYAR 237
Qy 233 AAAAGIVDPAHGTGVANAGLVYFNGRLAMSDDLQVQITPVGDLKTVGRFDFDGOLES 292
Db 238 GLFGLVDHSGKTGVANAGLVYFNNRLAMSDDLQVYHVKVPTGDLTKTEGDFDGOLES 297
Qy 293 TWIAHPKVDPSGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQDQPTMMHDFAIT 352
Db 298 TWIAHPKLDPSGELFALSYDVVSKPYLKYFRFSGKNGESNDVBIPEVPTMMHDFAIT 357
Qy 353 NPVVVPDQOVNFKLPEMIRGGSPVYDKNKVARFGLDKYADSSNIKWIDAPDCCFCHL 412
Db 358 NPVVVPDQOVNFKLPEMIRGGSPVYDKNKVARFGLDKYADSSNIKWIDAPDCCFCHL 417
Qy 413 NNAAEETDETVVIGSCMTPTDPSIFNESDENLKSILRLNLKTGSTRRPIISNEDQ 472
Db 418 NNAAEETDETVVIGSCMTPTDPSIFNECDGLKSVLSILRLNLKTGSTRRPIISNEDQ 477
Qy 473 QVNLBAGVNRNMLGKTKFAYLALAEAPKPVSGFAKVDLTGTVKHLVGNRYGGSPFL 532
Db 478 QVNLBAGVNRNMLGKTKFAYLALAEAPKPVSGFAKVDLTGTVKHLVGNRYGGSPFL 537
Qy 533 FLPFGE--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYGFHGTG 590
Db 538 FLPFDPNSKEEDDGVILAFVHDEKTKWSELQIVNAVSLVEATVKLPSRPVYGFHGTG 597
Qy 591 ADDLAKQ 597
Db 598 ANDLANQ 604

RESULT 3

US-09-758-269-12

; Sequence 12, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; TITLE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; CURRENT FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 12

; LENGTH: 612

; TYPE: PRT

; ORGANISM: Vigna unguiculata

US-09-758-269-12

Query Match 68.8%; Score 2168.5; DB 10; Length 612;
Best Local Similarity 68.6%; Pred. No. 1.2e-209;
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

Qy 8 AAVSGWLGNGHTQPPPLSSQSSDLSYCS---SLPWASVTRKLVNVSALHTPPALHPPKQ- 65
Db 3 SSAENWFWNLTPSPFPKLPDSTS-SPTNLLPL--RKTSSNITITCSLOT---LHPKQY 56
Qy 66 -----SSNSPAIVVVKPKAKESNT-----KQNLQFORAAAALDAA 100
Db 57 QPTSTSTSTATTTPITPIKTITTTTTTPPRETNPLSDTNQPLPQKWNFLQKAAATALDV 116
Qy 101 EGLVSHKHLPLPKTADPSVQIAGNFAVPNEQVRNLPVVGKLPDSIKGVYVRNGANP 160
Db 117 ETALVSHKHLPLPKTADPRVQIAGNFAVPVPEHAADQGLPVVGKIPKCIDGVYVRNGANP 176

192 CRFTETARLRQERAIGRVFPFKATGELHGHSGIARLALFYARAAACGLVDP SAGTVANAG 251
251 LAVFENGILLAMSEDDLPVQVQITPENGDLKTGREDPDGQLESTMIAPKPKVDPSGELFAL 310
252 LVFENGILLAMSEDDLPVHURVADDGULETVGRVDFDQGLGCAMIAHPKLDPATGELHAL 311
311 SYDVVSKPYLYKFRFSPDGTSPVETQLDQPTMMHDFAITENFVVVDPDQVVKFPEMI 370
312 SYDVVKRPYLYKFRFSPDGTSPVETQLDQPTMMHDFAITENFVVVDPDQVVKFPEMI 371
371 RGGSPVVVVKVAFGLDKYAEEDSSNIKWIDAPDCFCFHLNANAMEEPEETDE 430
372 RGGSPVVVLDKERTGRFVLPKHAADASEMAWVDPDCFCFHLNANAMEEPEETDE 431
431 MTPDPSIFNESDENLKSVLSEIRLNLTGSTRPPIISNEDQVNLNLEAGVNRNMLGRKT 490
432 MTPADPSIFNESDERLESVLTEIRLDARTGRSTRVAVLP-PSQENLEVGMVNRNMLGRES 490
491 KFAYLALAEAPKPKYSGFAKDLTTGCVKHHLYGDNRYGGEPLFLPGEGBE 545
491 RYAYLAVAEAPKPKYSGFAKDLTTGCVKHHLYGDNRYGGEPLFLPGEGBE 550
546 ILFCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAQV 597
551 VLTFFVHDERAGTSELLVYVNAADIRLEATVQLPSRVPGFHGTFTGQLEBAQ 602

RESULT 6

US-09-758-269-10
; Sequence 10, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-10

Query Match 52.8%; Score 1663.5; DB 10; Length 577;
Best Local Similarity 55.4%; Pred. No. 1.1e-158;
Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;
24 LSSQSDLSYCSSLPMSRVTRKLVSS-ALHTPALHPFKQSSNSPAIVVVKPKAKESN 82
9 LIPTKTSRSHLLPQPKNANISRRILINPFKIPTLPTDLSVPVSP-----VCLKPTYPN 62
83 TKOMNLFQRAAAALDAAE-GFLVSHKELHPLPKTADPSVQIAGNFAPVNEQVVRNLPV 141
63 ---LNLQKLATMLDKIESSIVPMEQNRPLPKPTDPAVLQSGNFAPVNECPVQNGLEV 119
142 VGKLPDSIKGVYVRNANPLHEPVTHHFFDGGGMVHAK--FEHGSASYACRFTQTRNF 199
120 VGOIPSLKGVYVRNANPMFPPLAGHLFDGGMTHAVS1GFDN-QVSSCRYTKTNRL 178
200 VQERQLGRPVFPKAIAGELHGHGTGIRMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLL 259
179 VQETALGRSVFPKPIGELHGHSGIARLALFTABAGIGLVDTGRGCVANAGVVFNGRLL 238

260 AMSDEDDLPYQVQITPENGDLKTGREDPDGQLESTMIAPKPKVDPSGELFALSYDVVSKPY 319
239 AMSDEDDLPYQVQITPENGDLKTGREDPDGQLESTMIAPKPKVDPSGELFALSYDVVSKPY 298
320 LKYFRFSPDGTSPVETQLDQPTMMHDFAITENFVVVDPDQVVKFPEMIRGSGSPVYD 379
299 LAYLKFNCTCGKTRDVEITLPEPTMIHDFAITENFVVVDPDQVVKFPEMIRGSGSPVYV 358
380 KUKVARFGLDKYAEEDSSNIKWIDAPDCFCFHLNANAMEEPEETDE-----VVVIGSCMTPPD 435
359 KEKMARFGLVSKQDLTGSNDINWVDPDCFCFHLNANAMEE-RTBEGDPPVIVVIGSCMTSPD 417
436 STFNESDENLKSVLSEIRLNLTGSTRPPIISNEDQVNLNLEAGVNRNMLGRKT KPAYL 495
418 TIFSESGETRVLSEIRLNMTKESNEKVIWVG----VNLEAGHINRSYVGRKSQFYVI 473
496 ALAEAPKPKYSGFAKDLTTGCVKHHLYGDNRYGGEPLFLPGEGBEEDGTYLFCFVHDEKT 555
474 AIADPWPCKSGIAKVDIQTGTVSEFNYGSPRFGGEPFCFVPEGEEDKGYVNGFVRDEEK 533
556 WKSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAQV 598
534 DESEFVVVDATDMKQVAAVRLPERVYPYGFHGTFFVSENQLEQV 576

RESULT 7

US-09-758-269-4
; Sequence 4, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-4

Query Match 30.7%; Score 968; DB 10; Length 595;
Best Local Similarity 36.7%; Pred. No. 2.2e-88;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;
42 SRVTRKLVSSALHTPALHPFKQSSNSPAIVVVKPKAKESNTKQMLFORAAAAALDAAE 101
42 SPITNPSDNDRENKPKTLH---NRTNHTLVSPPKLREPMTLATLAF---TTVEDVIN 94
102 GFLVSHKELHPLPKTADPSVQIAGNFAPVNEQVVRNLPVW-GKLPDSIKGVYVRNANP 160
95 TFDIP-----PSRPSVDPKHLVSDNPAVLDELPTDCEIIHGTLPLSLNGAVIRGNPNP 149
161 LHEFVTHHFFDGGGMVHAKVFEHGSASYACRFTQTRNFVQERQLGRPVFPKAIAGELHGH 220
150 QFLPRGPHYHLFDGGMHLAKIHNGKATLCRSRVKTKYNVEKQGTGAPVNPVFSGNGV 209
221 T-GTARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVQITPENGDLK 279
210 TASVARGALTAARVLTGQYVNVNIGLANTSLAFFSNRFLPAGESDLPFAVRLTESGDIE 269
280 TVGRFDFDQLESTMIAPKPKVDPSGELFALSYDVVSKPYLYKFRFSPDGTSPVETQLDQPTMMHDFAITENFVVVDPDQVVKFPEMI 338
270 TIGRYDFDGLKAMSMTAHPKTDITGBTFAFRYGPV--PPFLTYFRFDSAGKQORDVPYIFS 328

Db 344 EDEVLITCRLENPDLDWMSGVKKELENFNGNELYEMRNMKTGSQKLSASA----- 398
QY 476 LEAGWNRNMLGRKTFAYLALAEPPKVSFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDPRINECYTGKQRYVGTILDSIAKVTGIKFDLHAEAEFGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPEGGEDEGYILCFVHDEKTKWKSBLQIVNAVSLVE--ATVKLPSRV 580
Db 459 GEGRYGSEAIYVPRETAEDDGYLIFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGPHGTFIGADDLAKQVV 599
Db 519 PYGFHALFVTEEQLEQTL 537

RESULT 10
US-09-758-269-33
; Sequence 33, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-33

Query Match 29.7%; Score 936; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 3.2e-85;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIVVKPK-AKESNTKQMLFORAAAAALDAEGFLVSGHEKHLPLKPTADPSVQ 122
Db 4 KLSDCSIIISVHPRPSKGFSSKLLDLRLVVKLM-----HDSASLPLH-----Y 47
QY 123 TAGNPAFV-NEOPVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMHAKV 181
Db 48 LSGNPAFIRDETTPVKDLPHVHGFLECLNGEFVRVGNPKFDVAGVHWFDDGGMIRGVR 107
QY 182 FEHGSASACRTQTNRFVQERQLGRFPFKAIGELHGHGTGIARLMLFYAARAAGIVDPA 241
Db 108 IKDGKATYVSRYVTKSRKQEEFFGAAKFMK-IGDLKGFGLLVVQQLRKLKILDNT 166
QY 242 HGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNGLDKTVGRFPDGOLESTMAHPKVD 301
Db 167 YNGTANTALVYHGGKLLALQADKPYVVKLEDDGLQTLGIIDYDKELTHSFTAHKPKVD 226
QY 302 PSBSGLFALSVDWSKPYLYKFRFSPDGTSPDVEIQDQTMHDFAITENFVVPDQ 361
Db 227 PYTGEMFTFGYS-HTPPYLYTYRIVSKDGMHDPVPIITSEPMHDFAITETIYALFMDLP 285
QY 362 VYFKLPEMIRGGSPPV-YDNKVARFGILDKVAEDSSNIKVIDAPDCFCFLHNAWEPE 420
Db 286 MFRPRKEMVKEKMTYSDPTTKARFGVLPYAKDELIRWFELENCFIFFNANAW--E 343
QY 421 TDEVVVIGSCMTPD-----SIFNESDENLKSVLSEIRNLKGTSTRRPIISNEDQVN 475
Db 344 EDEVLITCRLENPDLDWMSGVKKELENFNGNELYEMRNMKTGSQKLSASA----- 398

QY 476 LEAGWNRNMLGRKTFAYLALAEPPKVSFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDPRINECYTGKQRYVGTILDSIAKVTGIKFDLHAEAEFGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPEGGEDEGYILCFVHDEKTKWKSBLQIVNAVSLVE--ATVKLPSRV 580
Db 459 GEGRYGSEAIYVPRETAEDDGYLIFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGPHGTFIGADDLAKQVV 599
Db 519 PYGFHALFVTEEQLEQTL 537

RESULT 11
US-08-976-063C-22
; Sequence 22, Application US/08976063C
; Publication No. US20020182697A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONIFERYL ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN
; TITLE OF INVENTION: ACID AND THEIR USE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTRA
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,063C
; FILING DATE: 21-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 196 49 655.1 (Germany)
; FILING DATE: 29-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-063C-22

Query Match 13.1%; Score 411.5; DB 8; Length 505;
Best Local Similarity 25.8%; Pred. No. 2.8e-32;
Matches 137; Conservative 79; Mismatches 192; Indels 123; Gaps 19;

QY 118 DPSVQIAGNFAPVNEQVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMV 177
Db 8 DP--QLVGTLLPRTIADLFLEVDGEI PKSI NGT FYRNTPEPVTPOKFTFIDGDGMA 65
QY 178 HAVKFEHGSASACRTQTNRFVQERQLGRFPFKAIGELHGHGTGIARLMLFYAARAAGI 237
Db 66 SAFPHFEDGHVDFISRWVKTFARFABELARKSLFGMYRNPYTDTSVKGL----- 114
QY 238 VDPAGHTGVANAGLVFNGRLLAMSEDDLPYQVQITPNGLDKTVGRFPDGOLES-TWIA 296
Db 115 -----DRTVANTSIISHHGKVLAVKEDGLPY--ELDP-PLTRGHFVDGQVTSQHTA 166

```

QY 297 HPKVDPSGELFALSVDVSKPYLKYFRFSPDGTGKSPDV-----EIQDOP- 342
Db 167 HPKYDPTGDL-----FFGSAKGAETPDMAYYIVDKHGKWTHTWEQPY 213
QY 343 -TMMHDFAITENFVVDPDQVVKPEMIRGGSPV-VYDKNKVARFGLDKYAESSNIK 400
Db 214 GAFMHDFAITENFVVDPDQVVKPEMIRGGSPV-VYDKNKVARFGLDKYAESSNIK 400
QY 401 WIDAPDCFCFHLNNAWEPETDEVVVIGSCWTP---PDS-----I 437
Db 271 WLKAPALWVFRVNVNAGVTKYIDLMSEILPFPFNPSONQFAPEKAVRTRWEIDL 330
QY 438 FNESEDNLKSVLSIRNLKTGST-----RRPIISNEDQOVNLEAGMV 481
Db 331 DSSDEIKRTHLHFFAEPMIMDSSPALQCNRYGFMGVDDPRKPLAHQAQAEKI----- 393
QY 482 NRMMLGKTKFAYLALABPKVSGFAKVDLTGCEVKKHLVGNRYGGEPLFLP-GEQGE 540
Db 384 -----FAYNSLG-1W-----DNHRGDYDLWYSGEASAAQEPFVPSGTAA 423
QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSL-VEATVKLPSRVPYGPFGTPI 589
Db 424 EGDGYLLTVVGRLDENRSDVLDTQDIQSGPVATIKLPLRLAALHGCWV 474

```

RESULT 12

US-10-149-759-66
 ; Sequence 66, Application US/10149759
 ; Publication No. US20030157592A1

GENERAL INFORMATION:

```

; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cippus, Petra
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins
; TITLE OF INVENTION: involved in the synthesis of tocopherols and
; TITLE OF INVENTION: carotenoids.
; FILE REFERENCE: BASF/NAE 1333/99 PCT/US
; CURRENT APPLICATION NUMBER: US/10149759
; PRIOR FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: PCT/EP/00/12698
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 66
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-10-149-759-66

```

```

Query Match      8.9%; Score 279; DB 12; Length 177;
Best Local Similarity 34.8%; Pred. No. 1.2e-19;
Matches 62; Conservative 34; Mismatches 66; Indels 14; Gaps 3;

QY 341 OPTMMHDFAITENFVVDPDQVVKPEMIRGGSPVVDKNKVARFGLDKYAESSNIK 400
Db 1 QIVMMHDFAITENYAFIWDLPILMDGESMMKGNFIKFDETKEARLGLVLPYATNESQLR 60
QY 401 WIDAPDCFCFHLNNAWEPETDEVVVIGSC-----MTPPDSIFNESDENLKSULSEIRL 454
Db 61 WFTIPVCFIFHNANAWEE---GDEIVLSRMEIEINLTAAADGFKENERISQPKLFEFRI 117
QY 455 NLKTGSTRRRPIISNEDQOVNLEAGMVNRNMLGKTKFAYLALABPKVSGFAKVDL 512
Db 118 NLKTGEVQKQL-----SVLVVDPPRVNVEYGRKTOYMGAINDKSSKMVGCKFDL 170

```

RESULT 13

US-10-053-192-1
 ; Sequence 1, Application US/10053192
 ; Publication No. US20030087336A1

```

; GENERAL INFORMATION:
; APPLICANT: BACHMANN, Heinrich
; APPLICANT: BRUGGER, Roland
; APPLICANT: FRIEDLEIN, Arno M
; APPLICANT: WIRTZ, Gabriele M
; APPLICANT: WOGGON, Wolf-Dietrich
; APPLICANT: WYSS, Adrian
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
; FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES,....
; CURRENT APPLICATION NUMBER: US/10/053,192
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 103382.0
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 526
; TYPE: PRT
; ORGANISM: CHICKEN
US-10-053-192-1

```

```

Query Match      8.3%; Score 260; DB 15; Length 526;
Best Local Similarity 24.4%; Pred. No. 6.1e-17;
Matches 133; Conservative 85; Mismatches 217; Indels 110; Gaps 29;

QY 132 EQVVRNLPVVGKLPDSIKGVVVRNGANPLHE--PVTGHHFFDGDGMVHVKFEHGSASY 189
Db 11 EHPEPIKAEVQGLPTWLGQVLLRNGPG-MHTIGTKYNHWFGLALLHSFTFKNGEVVY 69
QY 190 ACRF-----TOTNRFVQERQLGRPVFPAKAGELGHTGTCIARLMLFYARAAAGIVDP 240
Db 70 RSKYLRSDTYNCNIEANRIVS-BEGTWAYDDPCNKI-----FAKAFYLSHTTIDEFTDN 123
QY 241 AHTGTGANAGLVYFNGRLLAMSEDDLPYQVOITPNGDLKTVGRFDFDQGLSESTM-IAHPK 299
Db 124 CL-INIMTKDDY-----ATSETN--FIRKIDPQ-TLETLDKVDYSKVVAVNLATSHPH 174
QY 300 VDPESGELFALSVDVSKPYLKYFRFS-----PDGTSKPD-----VEI-----QLDQPT 343
Db 175 YD-SAGNLTNMGTSIVDKGRTKYLFKIPSSVPEKEKKSCFKHLEVVCSPSRSLQPS 233
QY 344 MMHDFAITENFVVDPDQVVKFL-----PEMIRG---GSPVVYDKNKVARFGLDKYAE 394
Db 234 YHSGEITENYIVFIEQP--FKLDIVKLATAVIRGVNNAWASCLSPHKEKDTWTFHVDKTK 291
QY 395 DSSNIK-WIDAPDCFCFHLNNAWEPETDEVVVIGSCWTPPDSIFN-----ESD 442
Db 292 KEVSTKFTYDA--LVLYHHINAYEE---DGHVVPDIVAYRDNLSYDMFYLKLDKDFEVN 346
QY 443 ENLKSVL-----SEIRLNL-----KTGSTRRRPIISNEDQOVN 476
Db 347 NKLTSTPTCKRFVFPLOYDKDAEVSNLVKLPTSATAVKEKDGSIYCOPELICE---GI 402
QY 477 EAGMVNRNMLGKTKFAYLALABPKVSGFAKVDLTGCEVKKHLVGNRYGGEPLFLPG 536
Db 403 ELPRVNYDNGKKYKVVATEVQWSPVPTKIAKLVQTKEV-LHWGEDHCWSPSEIFVPS 461
QY 537 -EGGEDEGVIL-CFVHDEKTKWSELQIVNAVSLV--EATVKLPSRVPYGPFGTFIGAD 592
Db 462 PDAREDEGVILTCVVVSEPNKAPFLLLIDAKTFKELGRATVNVEMHL--DLHGMIPON 519
QY 593 DLAKQ 597
Db 520 DLGAE 524

```

RESULT 14

US-10-053-192-4
 ; Sequence 4, Application US/10053192
 ; Publication No. US20030087336A1
 ; GENERAL INFORMATION:

APPLICANT: BACHMANN, Heinrich
APPLICANT: BRUGGER, Roland
APPLICANT: FRIEDLEIN, Arno M
APPLICANT: WIRTZ, Gabriele M
APPLICANT: WOGGON, Wolf-Dietrich
APPLICANT: WYSS, Adrian
APPLICANT: WYSS, Markus
TITLE OF INVENTION: BETA, BETA-CAROTENE 15, 15'-DIOXYGENASES, NUCLEIC ACID
FILE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
FILE REFERENCE: B-B-CAROTENE 15, 15'-DIOXYGENASES, ...
CURRENT APPLICATION NUMBER: US/10/053.192
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 103382.0
PRIOR FILING DATE: 1999-02-22
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 506
TYPE: PRT
ORGANISM: CHICKEN
US-10-053-192-4

Query Match 7.8%; Score 247; DB 15; Length 506;
Best Local Similarity 24.3%; Pred. No. 1.28-15;
Matches 130; Conservative 83; Mismatches 213; Indels 110; Gaps 29;
QY 132 EQVARNLPVVGKLPDSIKGVYVRNGANPLHE--PVTGHHFFDGDGMVHAVKPEHGSASY 189
DB 2 EHPPIKAEVQGLPTWLGVLRLNPG-MHTIGDTKYNHWFGLALLHSFTPKNGEVY 60
QY 190 ACRF-----TOTNRFVQERQLGRVFPKAGELHGHGTGIARLMLFYARAAAGIVDP 240
DB 61 RSKYLSRDTYCNIBANRIVVS-EFGTWAYDPCKNI-----FAKAFSYLSHTIPEFTDN 114
QY 241 AHGTGVANAGLVYFNGRLAMSEDDLPYQVITPNDGLKTVGRFPDQGLESTW-IAHPK 299
DB 115 CL-INIMKTGDY-----ATSEIN--FIRKIDPQ-TLETLDKYDYKYVAVNLTATSPH 165
QY 300 VDPESGELFALSYDVVSKYKIFRFS-----PDGTSKD-----VEI-----QLDQPT 343
DB 166 YD-SAGINLMGTSTVDXGRTKYVLFKIPSSVPEKEKKSCFKHLEVVCSPISPSRLQPS 224
QY 344 MMHDFAITENFVVDPQVWFKL-----PEMIRG-----GSPVYDKNVARFGILDKYAE 394
DB 225 YHSGFIGHTENYVITEQ--FKLDIVKLATAYIRGVNWKASCLSFHKEDKTFHFVDRKTK 282
QY 395 DSSNIX-WIDAPDCFCFLMNAWEPETDEVVIGSCWTPPDSIFN-----ESD 442
DB 283 KEVSTKFTYDA--LVLYHHINAYEE--DGHVVPDIVAYRDNLSLYDMFYLLKKLDKDEVN 337
QY 443 ENLKSVL-----SEIRLN-----KTGSTRPPIISNEDQVNL 476
DB 338 NKLTSTPTCKRFVPLQVDKREVGSLNVLKLETSATAVKEKDGSIYCOPELCE-----GI 393
QY 477 EAGMVRNMLGRKTFAYLALAEPPKVSFGFAKVDLTGTEVKKHLYGDNRYGGPFLPLPG 536
DB 394 ELPRVNYDYGKKYKYVATEVQWSPVPTKIAKLVQTKV-LHMGEDHCWSPSEPIFVPS 452
QY 537 -EGGDEGEYIL-CFVHDEKTKWSELOIVNAVSLV--EATVKLPSRVPYGHGTF 588
DB 453 PDAREDEGVWLTCCVVVSEPNKAPFLILLDAKTFKELGRATVNVENHL--DLHGMF 506

RESULT 15

US-10-168-517-19
Sequence 19, Application US/10168517
Publication No. US20030166595A1
GENERAL INFORMATION:
APPLICANT: Von Lintig, J.
APPLICANT: Vogt, K.
TITLE OF INVENTION: No. US20030166595A1 dioxigenases catalyzing cleavage of beta-ca
FILE REFERENCE: S-60094USPCT
CURRENT APPLICATION NUMBER: US/10/168, 517

CURRENT FILING DATE: 2002-06-19
PRIOR APPLICATION NUMBER: 00105822.1
PRIOR FILING DATE: 2000-03-20
PRIOR APPLICATION NUMBER: 99125895.5
PRIOR FILING DATE: 1999-12-24
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 19
LENGTH: 549
TYPE: PRT
ORGANISM: Danio rerio
US-10-168-517-19
Query Match 7.6%; Score 238.5; DB 12; Length 549;
Best Local Similarity 21.7%; Pred. No. 9.7e-15;
Matches 118; Conservative 89; Mismatches 219; Indels 119; Gaps 25;
QY 130 VNEQVVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGH-HFFDGDGMVHAVKPEHGSAS 188
DB 34 VEEIPDPITTLIKGOIPSWINGSFLRNGPFGKFEFGESKTFHWFDGMLMHRFNKDGQVT 93
QY 189 YACRFTQTNRFVQERQLGRVFPKAGELHGHGTGIARLMLFYARAAAGIVDPAHGTGVAN 248
DB 94 YSRRLQSDSYVQSEKNRIVVSE-FGTLATPDPCNI--FARFERSFOIP-KTTDNAG 148
QY 249 AGLVTFNGELLAMSEDDLPYQVITPNDGLKTVGRFDFDGOLE-STMIHAKVDESGEL 307
DB 149 VNFVYKGDFFVYSTETN--FMKIDP-VSLETKEKVDWSKFIASAAATAHPHYDRE-GAT 204
QY 308 FALSYDVVSKPYLKYPFRFSPDGTSKD-----VEIQLD-----OPTMHEDFAITENF 354
DB 205 YMGNSYGRKGFYHILRVPPGKQDDADLSGAILCSIPAADPRKPSYVHSFVMSY 264
QY 355 VVVPDQVVKLPPEM-----IRGGS---PVVDKNKVARFGILDKYAESSNIKWIDAPDC 407
DB 265 IVFIEQPIKLDLLKFLMYRIAGSKPHKVMWNPELDTIFHVADRHTGQLLNTKYYSA-M 323
QY 408 FCFHLMNAWEPETDEVVIGSCWTPPDSIFNE-SDENLKSVLSE-----RYGG 529
DB 324 FALHQNAYEE--NGYLIMDCGDDGNVIGEFLENLQSTGEDLDKFFNSLCTNLP 380
QY 452 --IRNLKTGSTRPPIIS-----NED--QOVNLEAGMWN-R 483
DB 381 VYLPLEVKEDENPDQNLINLPYTTASAVKTQTVGFVLYHEDLYNDDLLQYGGLEFPQINYA 440
QY 484 NMLGRKTFAYLALAEPPKVSFGFAKVDLTGTEVKKHLYGDN-----HYFGDLSLLKMDLEGKLVWRHAG 480
DB 441 NYNARPYRIFY-----ACGFG-----LVFGDLSLLKMDLEGKLVWRHAG 480
QY 530 ---EPLFLPG-EGGDEGEYILCFVHDEKTKWSE-LOIVNAVSLVLEATVKLPSRVPY 583
DB 481 LPSPVPVFIAPDAQDEDDGVVMSVITPREKKSGFLLVLDKATFTELGRABVPVDPY 540
QY 584 FHGTF 588
DB 541 THGLF 545

Search completed: November 14, 2003, 21:23:16
Job time : 38 secs

Db 442 FVPDRCKLVKTKETWVQEPDSYSEPIFVSHPDALBEDDGVLSVVVSPGAGQKPAY 501
QY 560 LQIVNAVSLEVEATVKLPSRVPYGFHGTFF 588
Db 502 LLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 4
US-08-488-305A-6
; Sequence 6, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vineet
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-488-305A-6

Query Match 6.6%; Score 209; DB 1; Length 533;
Best Local Similarity 20.7%; Pred. No. 8.7e-13;
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;

QY 105 VSHEKLHPLKPTADPSVQIAGNAPVNEQPVRRNLP-----VVGKLPDSIKGVVVRNGANP 160
Db 1 MSSQVEHP-----AGGYKKLPETVEJSSPLTAHTVGRPLWLTSGLRCGPG 49
QY 161 LH---EPVTGHHFDGGMHVAHFEGSASACRFTQTRNFVE-----RQLGRPV 209
Db 50 FEVGSERP--XHLDDGQALLHKFDKSGHVTHYHRRFRTDAYVRAMTEKIVITEGTCA 107
QY 210 PPKAIGELHGTGTLARMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLIAMSDE----- 264
Db 108 FPDFCKNI-----FSRFFSYF-----RGVEVTDNALV-----NIYFVGEDYYACT 147
QY 265 DLPYQVQITPNDLKTGRFDFDQGLE-STWIAHPKVDPS-----GELFALSVDV 314
Db 148 ETNFIKKNPE-TLETIKQVLDLNVSVNGATAPHIENDGTVYVNGICFGKFSIAYNI 206
QY 315 VSKPYLYKFRSPDGTSPDVEIQLD-----OPTMMHDFALTENFVVVDPQVVKLP 369
Db 207 VKIPPLQADKEDP--ISKSEIWWGPCSDREPKPSVHSFGLTPNYIVFVETPKVINLKF 264
QY 370 IRGGSPPVYDKNKVAF-----GILDKYAEADSSNIKWIDAPDCFCFHLWNAWEEP 419

Db 265 LSSWS--LWGANVMDCFESNETGMVHLIADKKKKVINKKYRTP--NFLFHINTYEDH 321
QY 420 ETDEVVIGSCMTTPDPSIFNESD--ENLKSVLSEIRLNLKGTGSTRPPIISNEDQOVLEA 478
Db 322 EF--LIVDLCCWKGFEFVINYLYLANLRENWEEVK--KNARKAPOPEVRYVLPINIDK 376
QY 479 GWNARNM--LGRKTKFAVLAALAEW--PKV--SG-----FAKVDLTT--GEVKKHLYG 523
Db 377 ADTGKNUVLTNTTATILCSDETIWLEPEVLFSGPQAEFFPOINVKYGGKPYTAYG 436
QY 524 -----DNRYGGEPLFLP--GEGGDEDEGYILCFVHDEKTV 556
Db 437 LGLNHFVDRCLKLVKTKETWVQEPDSYSEPIFVSHPDALBEDDGVLSVVVSPGAG 496
QY 557 K--SELQIVNAVSLEVEATVKLPSRVPYGFHGTFF 588
Db 497 QKPAYLLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 5
US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-30

Query Match 3.9%; Score 123.5; DB 3; Length 1043;
Best Local Similarity 20.6%; Pred. No. 0.0026;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TATAVSGRWLGGNHTQPPPLSSQSSDLSYCSSL--PMASRYTRK---LNVSSALHTPP 58
Db 153 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVYGEFGLLNPATGVMI 200

ATTORNEY/AGENT INFORMATION:

NAME: Verity, Hana
 REGISTRATION NUMBER: 30,518
 REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-324-1677
 TELEFAX: 650-324-1678
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1721 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-928-3613-6

Query Match 3.9%; Score 123.5; DB 3; Length 1721;
 Best Local Similarity 20.6%; Pred. No. 0.0064;
 Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;
 5 TATAVSGRWLGNNHTQPPLSOSSSDLSYCSSL--PMASRVTRK-----LNVSALHTPP 58
 831 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVVGEGFGLLNPAATGVMI 878
 59 ALHFPKQSSNSPAI-----VVKPKAKESNTKQNNL-----FQRAAAALD 98
 879 FLGPEQTFQSPFIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQKIDSISELMYD 938
 99 AAEGLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVNRNLPVVGKLPDPSIKGVYVRNGA 158
 939 IESGRLLGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPIDPTTGL----- 985
 159 NPLHEPVTGHFFPDGDMHVAHKEHGSASY--ACRFTQTNRFVQERQGRPVFPKAIGE 216
 986 -PFNPP-TGH-----LINPTNNMTDSSFAGAYKAVSNGIKTDNVYGLPV-----GE 1031
 217 LHGHTGIARLMFLYARAAAAGIVDPAHGTGVAN--AGLVYFNGL-LAMSEDDLPLYQVQIT 273
 1032 ITGLPKDPSDIPFNSTTGELVDPSTGKPINNSTAGIV--SGKPLPIEDE----- 1081
 274 PNGDLKTVGRFD-----FDQLESTMIHAKPKVDPESELFAISYDVVSKPKLYKFRFSP 327
 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNPETNSTVSGTSGTKP-----KP 1121
 328 -----DGTKSPDVEI--QLDQ-----PTMHDFAITENFVVVPDQVVF 364
 1122 GIPVNGGVVPDEAKQADKQKGLIVPPTNSINKDPVTNTQYSNTTGNIIINPETGKV- 1180
 365 KLPFEMIRG-----GSPVVYDNKVKARFGILDKYAEDSSNIKWIDAPDCFCFLMNAWE 417
 1181 -IPGSLPGSLNVPSEFNTPOQTDE-----ITGKPVDTVTGLPY----- 1216
 418 EPETDEVVVGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRPPIISNEDQQVNL 477
 1217 DPSTGEIIDPATKPIPGSVAGD-----EILTEV-LNITDEVVTGLPI-----DLE 1261
 478 AGVNRNMLGRKTKFAYLALAEPPKVSQFAKV---DLTTGEVKKHLYGDNRYGGEPLFL 534
 1262 TGL-----PRDPVSGLPQLNGTLVDPNSKKPIPGS--HSG--FI 1297
 535 PGEGEE 541
 1298 NGTSGEQ 1304

RESULT 8

US-09-588-995A-6
 Sequence 6, Application US/09588995A
 Patent No. 6514697
 GENERAL INFORMATION:
 APPLICANT: PETERSEN, CAROLYN
 APPLICANT: BARNES, DEBRA A.
 APPLICANT: NELSON, RICHARD C.

APPLICANT: GUT, JIRI
 TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
 TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
 TITLE OF INVENTION: INFECTIONS
 FILE REFERENCE: 480.19-5
 CURRENT APPLICATION NUMBER: US/09/588.995A
 CURRENT FILING DATE: 2000-06-06
 PRIOR APPLICATION NUMBER: 08/827,171
 PRIOR FILING DATE: 1997-03-27
 PRIOR APPLICATION NUMBER: 08/928,361
 PRIOR FILING DATE: 1997-09-12
 PRIOR APPLICATION NUMBER: 08/700,651
 PRIOR FILING DATE: 1996-08-14
 PRIOR APPLICATION NUMBER: 08/415,751
 PRIOR FILING DATE: 1995-04-03
 NUMBER OF SEQ ID NOS: 115
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 6
 TYPE: PRT
 LENGTH: 1721
 ORGANISM: Cryptosporidium parvum
 US-09-588-995A-6

Query Match 3.9%; Score 123.5; DB 4; Length 1721;
 Best Local Similarity 20.6%; Pred. No. 0.0064;
 Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;
 5 TATAVSGRWLGNNHTQPPLSOSSSDLSYCSSL--PMASRVTRK-----LNVSALHTPP 58
 831 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVVGEGFGLLNPAATGVMI 878
 59 ALHFPKQSSNSPAI-----VVKPKAKESNTKQNNL-----FQRAAAALD 98
 879 FLGPEQTFQSPFIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQKIDSISELMYD 938
 99 AAEGLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVNRNLPVVGKLPDPSIKGVYVRNGA 158
 939 IESGRLLGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPIDPTTGL----- 985
 159 NPLHEPVTGHFFPDGDMHVAHKEHGSASY--ACRFTQTNRFVQERQGRPVFPKAIGE 216
 986 -PFNPP-TGH-----LINPTNNMTDSSFAGAYKAVSNGIKTDNVYGLPV-----GE 1031
 217 LHGHTGIARLMFLYARAAAAGIVDPAHGTGVAN--AGLVYFNGL-LAMSEDDLPLYQVQIT 273
 1032 ITGLPKDPSDIPFNSTTGELVDPSTGKPINNSTAGIV--SGKPLPIEDE----- 1081
 274 PNGDLKTVGRFD-----FDQLESTMIHAKPKVDPESELFAISYDVVSKPKLYKFRFSP 327
 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNPETNSTVSGTSGTKP-----KP 1121
 328 -----DGTKSPDVEI--QLDQ-----PTMHDFAITENFVVVPDQVVF 364
 1122 GIPVNGGVVPDEAKQADKQKGLIVPPTNSINKDPVTNTQYSNTTGNIIINPETGKV- 1180
 365 KLPFEMIRG-----GSPVVYDNKVKARFGILDKYAEDSSNIKWIDAPDCFCFLMNAWE 417
 1181 -IPGSLPGSLNVPSEFNTPOQTDE-----ITGKPVDTVTGLPY----- 1216
 418 EPETDEVVVGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRPPIISNEDQQVNL 477
 1217 DPSTGEIIDPATKPIPGSVAGD-----EILTEV-LNITDEVVTGLPI-----DLE 1261
 478 AGVNRNMLGRKTKFAYLALAEPPKVSQFAKV---DLTTGEVKKHLYGDNRYGGEPLFL 534
 1262 TGL-----PRDPVSGLPQLNGTLVDPNSKKPIPGS--HSG--FI 1297
 535 PGEGEE 541
 1298 NGTSGEQ 1304

RESULT 9

US-08-928-361B-11
; Sequence 11, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1042 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-11

Query Match 3.7%; Score 116.5; DB 3; Length 1042;
Best Local Similarity 19.9%; Pred. No. 0.014;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHFPKQSSNGPAI-----VVKPKAKESNTKOMNL----- 88
Db 190 LNPATGVMIPGSLGPSEQTFPFEIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQ 249
QY 89 -FORAAAAALDAAGFLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVVRNLPVVGKLPD 147
Db 250 KIDSISELMYDIESGRLLIGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPID 302
QY 148 SIKGVYVRGANPLHPEVTHGHFFDGDGMVHAKVEHGSASY--ACRFTOTNRFVQERQL 205
Db 303 PTTGL-----PFNPP-TGH-----LINPTNNTMSSFAGAYKAVSNGIKTDNVY 347
QY 206 GRPV-----FPKAIHELHGTGIARLMLFYARAAAGIVDPAHGTGVAN-- 248
Db 348 GLPVDIEITGLPKDPVSDIPFNSITGEL-----VDPSTGKPINNYT 387
QY 249 AGLYVFNGRLLAMSEDDLQVQVITPNGDL---KTVGRPDFDGOLESTMIAPKVPDPESG 305
Db 388 AGIV-SGKRGLPPIEDE-----NGNLFDPSTKLPIDGNQL-----VNPEIN 428
QY 306 ELFALSVDVVSFKYLKYFRFSP-----DGTSPDVEI--QLDQ-----P 342
Db 429 STVSGTSGSTKE-----KPGIPVNGGVPEEAKQADKGDGLIVPPTNSIMKDP 481

QY 343 TMMHDFALTENFVVVDQOVVFKLPEMIRG-----GSPVVYDKNKVAREFGILDKYAE 395
Db 482 VTNTQISNTGNIINPFTGKV--IPGSLPGSLNVPFNTPOCTDE-----ITGKPVDT 532
QY 396 SSNIKWIDAPDCFCFHLNNAWEEPETDEVVVGSCMTPTDIFNESDENLKSIVSEIRLN 455
Db 533 VTGLPY-----DPSTGIIIDPATKLPPIGSAVD-----EILTEV-LN 569
QY 456 LKTGSTRRPLISNEDQOVNLEAGVMNRNMLGRKTKPAYLALABPWPVKVSGFAKV---DL 512
Db 570 ITTDEVTLGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 602
QY 513 TTGEVYKCHLYGDNRYGGBPLFLPGEGBE 541
Db 603 VDPSENKKPIPGS--HSG---FINGTSGEQ 626

RESULT 10
US-09-588-995A-11
; Sequence 11, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1042
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-11

Query Match 3.7%; Score 116.5; DB 4; Length 1042;
Best Local Similarity 19.9%; Pred. No. 0.014;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHFPKQSSNGPAI-----VVKPKAKESNTKOMNL----- 88
Db 190 LNPATGVMIPGSLGPSEQTFPFEIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQ 249
QY 89 -FORAAAAALDAAGFLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVVRNLPVVGKLPD 147
Db 250 KIDSISELMYDIESGRLLIGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPID 302
QY 148 SIKGVYVRGANPLHPEVTHGHFFDGDGMVHAKVEHGSASY--ACRFTOTNRFVQERQL 205
Db 303 PTTGL-----PFNPP-TGH-----LINPTNNTMSSFAGAYKAVSNGIKTDNVY 347
QY 206 GRPV-----FPKAIHELHGTGIARLMLFYARAAAGIVDPAHGTGVAN-- 248
Db 348 GLPVDIEITGLPKDPVSDIPFNSITGEL-----VDPSTGKPINNYT 387
QY 249 AGLYVFNGRLLAMSEDDLQVQVITPNGDL---KTVGRPDFDGOLESTMIAPKVPDPESG 305
Db 388 AGIV-SGKRGLPPIEDE-----NGNLFDPSTKLPIDGNQL-----VNPEIN 428
QY 306 ELFALSVDVVSFKYLKYFRFSP-----DGTSPDVEI--QLDQ-----P 342

Db 429 STVSGSTGSKP-----KPGIPVNGGVVDEEAKQADKQDGLIVPPTNSINKDP 481
QY 343 TMHEDFAITENFVVDPDQVQVFKLPEMIRG-----GSPVYDKNKVARFGILDKYABD 395
Db 482 VTNQVSNNTGNIINPETGKV--IPGSLFGSLNYPSPNTPOQTDE-----ITGKPVDT 532
QY 396 SSNIKIDAPDCFCFHLNABEPEDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLN 455
Db 533 VTGLPY-----DPSTGEIIDPATKLPFGSVAGD-----EILTEV-LN 569
QY 456 LKTGSTRAPITISNEDQVNLKAGVNRNMLGRKTKFAYLALAEWPVKVSGPAKV---DL 512
Db 570 ITTDEVTLGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 602
QY 513 TTGEVKKHLYGDNRYGGEPLFLPGEEGEE 541
Db 603 VDPSNKKPIPGS--HSG---FINGTSGEQ 626

RESULT 11

US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS, ANALOGS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-5

Query Match 3.7%; Score 116.5; DB 3; Length 1837;

Best Local Similarity 19.9%; Pred. No. 0.039;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHFFKQSNSEPAI-----VVKPKAKESNTKQNL-----88
Db 985 LNPATGWIPIGSLGPEQTPFSEIEDGGIIPPEVAANADKFKLSIPSPVESIPEDQ 1044

QY 89 -FQRAAAALDAEGLVSHSHKHLPLPKTADPVSQVQIAGNAPVNEQVRNLPVVGKLPD 147
Db 1045 KIDSISLMDYDIESGRLLIGQVSKRPIPGS-----IAGDLNPIPKTKTQTD-SYTGKPID 1097
QY 148 SIKGVYVRNGANPLHEHPTGTHHFPDGDGMVHAVKFEHGSASY--ACRFTOTNRFVQBRQL 205
Db 1098 PTTGL-----PNPP-TGH-----LINPINNTMDSFAGAYKYAVSNGIKTDNVY 1142
QY 206 GRPV-----PFKAIGELHGTGIAARMLFYARAAAGIIVDPAHGTGVAN-- 248
Db 1143 GLPVEITGLPKDFVSDIPFNSTTGL-----VDPSTGKPINNYT 1182
QY 249 AGLVYFNGRLLAXSEDDLPYQVQITPNGDL---KTVGRFPDGDQLESTMIAPKVPDESG 305
Db 1183 AGIV-SGKRGLPPIEDE-----NGNLFDSSTKLPIGNNQL-----VNPETN 1223
QY 306 ELFALSYDVVSKPYLYKFRFSP-----DGTKSDDVEI--QLDQ-----P 342
Db 1224 STVSGSTGSKP-----KPGIPVNGGVVDEEAKQADKQDGLIVPPTNSINKDP 1276
QY 343 TMHEDFAITENFVVDPDQVQVFKLPEMIRG-----GSPVYDKNKVARFGILDKYABD 395
Db 1277 VTNQVSNNTGNIINPETGKV--IPGSLFGSLNYPSPNTPOQTDE-----ITGKPVDT 1327
QY 396 SSNIKIDAPDCFCFHLNABEPEDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLN 455
Db 1328 VTGLPY-----DPSTGEIIDPATKLPFGSVAGD-----EILTEV-LN 1364
QY 456 LKTGSTRAPITISNEDQVNLKAGVNRNMLGRKTKFAYLALAEWPVKVSGPAKV---DL 512
Db 1365 ITTDEVTLGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 1397
QY 513 TTGEVKKHLYGDNRYGGEPLFLPGEEGEE 541
Db 1398 VDPSNKKPIPGS--HSG---FINGTSGEQ 1421

RESULT 12

US-09-588-995A-5
; Sequence 5, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1837
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-5

Query Match 3.7%; Score 116.5; DB 4; Length 1837;

Best Local Similarity 19.9%; Pred. No. 0.039;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSAHTPPALHPKQSSNPAL-----VVKPKAKSNTKQML----- 88
Db 985 LNPATGVMIPGSLGSETPSPEDGGIIPPEVAANADKFKLSIPSPVSEIPKQD 1044
QY 89 -FORAAAALDAAGFLVSHKHLPLKPTADPSQIAGNFAPVNEQVRNLPVVGKLPD 147
Db 1045 KIDSISELMYDIESGRLLIGQVSKRPISG-----IAGDLNPIWKTPTQTD-SWTGPID 1097
QY 148 SIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVFEHGSASY--ACRFTOTNRVQBRQL 205
Db 1098 PTTGL-----PENPP-TGH-----LINPTNNTWDSFAGAYKAVSNGIKTDNVY 1142
QY 206 GRPV-----FPAIGELHGHGIARLMLFYARAAGVDPHAGTGVAN-- 248
Db 1143 GLVDEITGLPKDPVSDIPFNSITTEL-----VDPSTGKPIINNYT 1182
QY 249 AGLVYFNGRLIAMSDDLPYQVQIIPNGDL---KTVGRFDPDGGQLESTMIAPKVPDPSG 305
Db 1183 AGIV-SGKGLPPIEDE-----NGNLPDPSTKLPIDGNQL-----VNPETN 1223
QY 306 ELFALSVDVVKPKLYKFRFSP-----DGTSPDVEI---QLDQ-----P 342
Db 1224 STVSGTSGSTKP-----KFGIPVNGGVVDPDEAKDQADKQGLIVPPTNSINKDP 1276
QY 343 TWHDFALITENVVVDPQVVKFLEPMIRG-----GSPVVYDKNKVAREGILDKYAED 395
Db 1277 VINTQVSNNTGNIINPFGKV--IFGSLPGSLNYPSPNTPOQTD-----ITGKPVDT 1327
QY 396 SNIKWIDAPDCCFHLWNAMEPETDEVVIGSCWTPPDSIFNEDENLKSVLSEIRLN 455
Db 1328 VTGLPY-----DPSTGEIIDPATKLPISGVAGD-----EILITEV-LN 1364
QY 456 LKTGSTRRPIISNEDQVNLKAGVNRNMLGRKTKFAYLALAPWPKVSGFAKV---DL 512
Db 1365 IITDEVGTGLPI-----DLETGL-----PRDPVSLPLQPNGL 1397
QY 513 TTGEVKKHLYGNRYGGBLPLFPGEGGEE 541
Db 1398 VDPNKKPIPGS--HSG---FINGTSGEQ 1421

RESULT 13
US-08-311-731A-403
; Sequence 403, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 403:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae
US-08-311-731A-403

Query Match 3.6%; Score 112; DB 4; Length 60;
Best Local Similarity 42.3%; Pred. No. 0.00028;
Matches 22; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 160 PLHEPVTGHHFFDGDGMVHAVFEHGSASYACRFTOTNRVQBRQLGRPVFP 211
Db 2 PLHPEFTVHPFDEDDGMLHMGVGRDGAFFRNSFRTDGLSENIAEPLWP 53

RESULT 14
US-09-461-697-77
; Sequence 77, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENET NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-77

Query Match 3.4%; Score 106; DB 3; Length 412;
Best Local Similarity 20.6%; Pred. No. 0.035;
Matches 74; Conservative 45; Mismatches 131; Indels 110; Gaps 16;

QY 23 PLSGSSDLSYCSLPMASRVTRKLVNSSLHTPPALHPFKQSSNSPAIVVKPAK--- 79
Db 9 PISSTQNRSSF-DLLPREFRLVE-----VHDEPP-LHQSANKPKPTMLDIPSEPCS 59
QY 80 -----ESNTKQMLFQRAAAALDAAEGFLYSHEKHLPLKPTADPSVOIAGNFAPV 130
Db 60 LTHHTQLQCHNRRLNLIATAQANQQQTEG--VKTESEPLP-SCPGSPPLDPLLFL 116
QY 131 N-EQPVVRNLPVVGKLPDSIKGVYVRNGANPLHEPVT-----GHHF 170
Db 117 DCKNP---NAPFQIRHSDPESDFYRGK-----EPVTELSWHSCROLLYQAVATILAHAG 168
QY 171 FDGDGMVHAVKFEHGSASYACRFTOTNRVQBRQ---LGRPVVPKALGELHGTGIARLML 228
Db 169 FDCANESVLETLTDVAHEYCLKFTKLLRFAVDEARLQTPFPDVMQVFEHVGIGSVLS 228
QY 229 F-----YARAAAGIVDPHAGTGVANAGLVYFNGRLIAME 263
Db 229 LQKFWQHRKIDYHSMYLMQISKQSEYER---IVNPEKAT-----E 266
QY 264 DDLPYQVQITPFGDLKTVGRFDFDGGQLESTMIAPKVPD-----ESGELFALSVDVVKSP 318

Db 267 DAKPVKIKBFPVSDI-----TFPVSEELADLASQSLPMGVILGAOSERPPSNLEVEAGP 322
US-09-107-532A-5705
RESULT 15
; Sequence 5705, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5705:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...471
; SEQUENCE DESCRIPTION: SEQ ID NO: 5705:
US-09-107-532A-5705
Query Match 3.3%; Score 105.5; DB 4; Length 471;
Best Local Similarity 19.7%; Pred. No. 0.05; Indels 175; Gaps 27;
Matches 97; Conservative 72; Mismatches 149;
Qy 179 AKFEHGSASVACFTTNR---PVQRQLG-----RPVFPKAI-----GELHGH 220
Db 47 SWIRFGDTTVCISATIENSVPFLRETGTGWTAEYSMLPRATSTNRRESSGKLFGR 106
Qy 221 T-GIARLMLFYARAAGIVPAHGTGVANAGLVYFNGELLAMSEDDLPHYOVQITPNDLK 279
Db 107 TMEIQRLI---GRSLRAVVD-----LEKLGERSIIVDCD-----VIQADGGTR 146
Qy 280 TVGRFDFDGLQLESTMIHPKV-----DPESGELFALSVDVVSFKYKYPFRSPDGTK 331
Db 147 TA---SITGAPVALKLAIEKLREKELSEDPIKEHLAAVSVGIL-----PDGTC 192
Qy 332 SPDVEIQLDQPTMMH-DFATEN--FVVV--PDQOVF---KLPEMIRGSGPVVYDNKV'383

Db 193 VTDLQYQEDSAALVDMNLVMTESGKFVEIQGTGBEATFDGSQLNEMLFFGKNAIEDLIKE 252
Qy 384 ARFGILDKYABDSSNIKWIDAPDCFCFHLNMAWBEPEDETVVVICSCMTPPD-----SIF 438
Db 253 QKHALLTEFAQNDERI-----SETKT-----IIATRNPGKAEFRNMF 291
Qy 439 NESDENLKSVL-----SEIRLNKLT-GESTRPLISNED-QQVNLKXGM 480
Db 292 KEAGYHVKTLLDYPELDPVEETGTSTFEENARKLAETIAQLDDQPVLAADDGLKVDALGGM 351
Qy 481 -----VNRNMLGRKTKPAY-----LALAEP-----WPK 503
Db 352 PGIIYSARFAGQKSDAGNNAKLLYELTDVPDEKTAQHCTLVFAAPKKDSLVVEAEP- 410
Qy 504 VSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGSGGEDEGYILCFVHDEKTKWSELOIV 563
Db 411 -----GRVARIPSENGFGYDLPFIP-EGKQTAAEL-----SSEK- 446
Qy 564 NAVSLEVEATVKL 576
Db 447 NKISHRAQAMKKL 459

Search completed: November 14, 2003, 21:18:46
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:11:32 ; Search time 47 Seconds
(without alignments)
2022.920 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGNGHT.....VPYGHGTFIGADDLAKQV 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	3150	100.0	599	22	AAE04784	Arabidopsis thalia
2	3150	100.0	599	23	ABR92311	Herbicidally activ
3	2280.5	72.4	605	22	AAE04789	Lycopersicon escul
4	2280.5	72.4	605	22	AAE04789	Neoxanthin cleavag
5	2168.5	68.8	612	22	AAE04787	Vigna unguiculata
6	2044.5	64.9	589	23	ABR91182	Herbicidally activ
7	1991	61.2	583	22	AAE04782	Arabidopsis thalia
8	1930	61.3	604	22	AAE04788	Zea mays neoxanthi
9	1930	61.3	604	22	AAE04788	Neoxanthin cleavag

10	1663.5	52.8	577	22	AAE04786	Arabidopsis thalia
11	1663.5	52.8	577	23	ABR92416	Herbicidally activ
12	968	30.7	595	21	AAG31333	Arabidopsis thalia
13	968	30.7	595	22	AAE04783	Arabidopsis thalia
14	968	30.7	595	22	AAE04783	Neoxanthin cleavag
15	959	30.4	517	21	AAG31334	Arabidopsis thalia
16	939	29.8	538	21	AAG31316	Arabidopsis thalia
17	939	29.8	538	21	AAG31315	Arabidopsis thalia
18	938	29.8	538	22	AAE04785	Arabidopsis thalia
19	937	29.7	538	22	AAE04790	Arabidopsis thalia
20	925.5	29.4	501	21	AAG31117	Arabidopsis thalia
21	857.5	27.2	544	22	AAE04783	Sunflower neoxanth
22	834.5	26.5	431	21	AAG31335	Arabidopsis thalia
23	745	23.7	446	22	AAE04783	Neoxanthin cleavag
24	322.5	10.2	570	22	AAE12086	Arabidopsis thalia
25	307.5	9.8	596	22	AAE12085	Arabidopsis thalia
26	279	8.9	177	22	AAE04781	Physcomitrella pat
27	272.5	8.7	166	21	AAG32792	Zea mays protein f
28	264.5	8.4	153	21	AAG32793	Zea mays protein f
29	260	8.3	526	21	AAE04781	Beta, Beta-caroten
30	259.5	8.2	152	21	AAG32794	Zea mays protein f
31	256	8.1	516	22	AAE04784	Amino acid sequenc
32	247	7.8	506	21	AAE04784	Beta, Beta-caroten
33	238.5	7.6	549	22	AAU04293	Zebra-2, beta-caro
34	232.5	7.4	532	22	AAU04292	Mouse-2, beta-caro
35	225	7.1	547	23	ABP63051	Human polypeptide
36	223	7.1	547	22	AAE04781	Human protein sequ
37	210	6.7	533	22	AAE04781	Canine RPE55. Can
38	209.5	6.7	539	22	AAE04781	Human protein sequ
39	209	6.6	533	14	AAE04781	Human retinol bind
40	207.5	6.6	556	22	AAU04294	Human-2, beta-caro
41	206.5	6.6	579	22	AAE04781	Human RECAP polye
42	185.5	5.9	529	21	AAE04781	Beta, Beta-caroten
43	154.5	4.9	65	23	ABP07713	Human ORFX protein
44	150	4.8	620	22	ABP07713	Drosophila melanog
45	150	4.8	620	22	AAU04291	Drosophila beta-ca

ALIGNMENTS

RESULT 1
AAE04784
ID AAE04784 standard; Protein; 599 AA.
XX
AC AAE04784;
XX
DT 10-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3.
XX
KW Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant.
XX
OS Arabidopsis thaliana.
XX
FN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PA (RIKE) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI; 2001-400081/43.
XX
N-PSDB; AAD09396.

PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance

Claim 3; Fig 9; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An acid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present sequence is Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein. The AtNCE3 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPRD5 (ComPea Responsive to dehydration) gene isolated from cowpea plant as a probe.

XX Sequence 599 AA;

Query Match	100.0%; Score 3150; DB 22; Length 599;
Best Local Similarity	100.0%; Pred. No. 28-310;
Matches 599; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy 1	MASFTATAAASGRMLGGNHTQPLPSSSQSDLSYCSSLPMASRVTRKLNVSALHTPPAL 60
Db 1	MASFTATAAASGRMLGGNHTQPLPSSSQSDLSYCSSLPMASRVTRKLNVSALHTPPAL 60
Qy 61	HPFKOSSNSPAIVYKPKAKESNTQKMLFORAAAAALDAAGFLVSHKELHPLKPTADPS 120
Db 61	HPFKOSSNSPAIVYKPKAKESNTQKMLFORAAAAALDAAGFLVSHKELHPLKPTADPS 120
Qy 121	VQIAGNFAPVNEQPVRRLNPVVGKLPSDIKGVYVRGANPLHEPVTGHHFFDGDGMVHAV 180
Db 121	VQIAGNFAPVNEQPVRRLNPVVGKLPSDIKGVYVRGANPLHEPVTGHHFFDGDGMVHAV 180
Qy 181	KFEHGSASVACRFQTQNRVFQEROLGRPVPKAIGELHGHGTGARLMLFYARAAAGIYDP 240
Db 181	KFEHGSASVACRFQTQNRVFQEROLGRPVPKAIGELHGHGTGARLMLFYARAAAGIYDP 240
Qy 241	AHGTGVANAGLVYFNGRLLAMSEDDLPIYQVQITPBGDLKTVGRFDPDGOLESTMTIAHPKV 300
Db 241	AHGTGVANAGLVYFNGRLLAMSEDDLPIYQVQITPBGDLKTVGRFDPDGOLESTMTIAHPKV 300
Qy 301	DPESGELFALSVDVVSKEPLYKFRFSPDGTKSPDVEIQLDQPTMHDFPAITENFVVVVDQ 360
Db 301	DPESGELFALSVDVVSKEPLYKFRFSPDGTKSPDVEIQLDQPTMHDFPAITENFVVVVDQ 360
Qy 361	QVVFKLPEMIRGSPVWYDKKVAARFGLDKYAEADSSNIKWIDAPPCFCFHLNWAESPE 420
Db 361	QVVFKLPEMIRGSPVWYDKKVAARFGLDKYAEADSSNIKWIDAPPCFCFHLNWAESPE 420
Qy 421	TDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLKTGESTRRPISNEDQQVNLEAGM 480
Db 421	TDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLKTGESTRRPISNEDQQVNLEAGM 480
Qy 481	VNRNMLGRKTKFAYLALAAPKPYKSGFAKVDLTTGVEVKKHLYGDNVYGESEPLFQEGGE 540
Db 481	VNRNMLGRKTKFAYLALAAPKPYKSGFAKVDLTTGVEVKKHLYGDNVYGESEPLFQEGGE 540
Qy 541	EDEGYILCFVHDEKTKWKGELQIVNAVSLVEEATVKLPSPVPYGFHGTTCGADDLAKQVV 599
Db 541	EDEGYILCFVHDEKTKWKGELQIVNAVSLVEEATVKLPSPVPYGFHGTTCGADDLAKQVV 599

RESULT 2
ABB92311

ID	ABB92311 standard; Protein; 599 AA.
XX	AC ABB92311;
XX	DT .31-MAY-2002 (first entry)
DE	Herbicidally active polypeptide SEQ ID NO 1522.
KW	Herbicidal; plant; agriculture; herbicide.
OS	Arabidopsis thaliana.
XX	WO200210210-A2.
XX	PD 07-FEB-2002.
XX	PF 28-AUG-2001; 2001WO-EP09892.
XX	PR 28-AUG-2001; 2001WO-EP09892.
XX	(FARB) BAYER AG.
XX	Tietjen K, Weidler M;
XX	WIPI; 2002-269010/31.
XX	Identifying plant target proteins for herbicidally active compounds,
PT	comprising aligning and comparing nucleic acid or amino acid sequences
PT	from plant with nucleic acid or amino acid sequences from non-plant
PT	organisms -
XX	Claim 5; SEQ ID NO 1522; 261pp + Sequence Listing; English.
XX	The invention relates to identifying target proteins
CC	(AB990790-AB994016) for herbicidally active compounds, comprising
CC	aligning and comparing nucleic acid or amino acid sequences from plant
CC	with nucleic acid or amino acid sequences from non-plant organisms using
CC	suitable search parameters, where plant sequences having an E-value
CC	greater by a factor of 3 than the E-value of most similar non-plant
CC	sequences are selected. The polypeptides or nucleic acids encoding them
CC	are useful for identifying modulators. The identified modulators are
XX	useful as herbicides.
XX	Sequence 599 AA;
SQ	Query Match 100.0%; Score 3150; DB 23; Length 599;
	Best Local Similarity 100.0%; Pred. No. 2e-310;
	Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MASFTATAAASGRWLGGNHTQPPLSSQSDDLSCSLPMASRVTRKLNVSALHTPPAL 60
Db	1 MASFTATAAASGRWLGGNHTQPPLSSQSDDLSCSLPMASRVTRKLNVSALHTPPAL 60
QY	61 HFPKOSSNSPAIVVKPKAKESNTKOMLFORAAAALDAAEGFLVSHKHLPLKTADPS 120
Db	61 HFPKOSSNSPAIVVKPKAKESNTKOMLFORAAAALDAAEGFLVSHKHLPLKTADPS 120
QY	121 VOIAGNFAPVNEQVRRENLPVVGKL PDSIKGVTVRNGANPLHEPVTGHHFFDGDGNHAV 180
Db	121 VOIAGNFAPVNEQVRRENLPVVGKL PDSIKGVTVRNGANPLHEPVTGHHFFDGDGNHAV 180
QY	181 KEFHGSASYACRFQTNRFWOERQLGRPVPFKAI GEHLHGHTGARLMLFYAAAAAGIVDP 240
Db	181 KEFHGSASYACRFQTNRFWOERQLGRPVPFKAI GEHLHGHTGARLMLFYAAAAAGI VDP 240
QY	241 ASGTGTVANAGLVYFNGLLAMSEDDLYEQVIITPNGDLKTVGRFDPDGQLESTMIAPHKV 300
Db	241 ASGTGTVANAGLVYFNGLLAMSEDDLYEQVIITPNGDLKTVGRFDPDGQLESTMIAPHKV 300
QY	301 DPESGEISALSYDVVSKPYLYKYFRFSPDGTKSDVDVFI QLDQPTMHEDFAITENFVVVPDQ 360
Db	301 DPESGEISALSYDVVSKPYLYKYFRFSPDGTKSDVDVFI QLDQPTMHEDFAITENFVVVPDQ 360


```

XX SQ Sequence 604 AA;
Query Match 61.3%; Score 1930; DB 22; Length 604;
Best Local Similarity 63.3%; Pred. No. 1.9e-186;
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSQSSDLSYCSLPMASRVTRKLNVSALHTP---PALHFPKQSSNSPAIVV---- 74
Db 17 PARSARAS-----NSVRFSPRAVSSVPPAECLQAPFHKPVADLPAPSRKPAIAVPGHA 71
QY 75 -KPKAKESNTKQNLKFORAAALDA-AEGFLVS-HEKHLPLKPTADPSVOIAGNAPVN 131
Db 72 AAPRKAEQKKQNLKFORAAALDAFEGFVANVLERPHGLPSTADPAVQIAGNAPVG 131
QY 132 EQVVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAKFEGHSA-SYA 190
Db 132 ERPPVHELPSGRIPFIDGVYARNGANPCFDPVAGHHLFDGDMVHALIRNGAAESYA 191
QY 191 CRFTQNRVFOERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDPAGTGVANAG 250
Db 192 CRFTETARLQERAIQRVFPKAIAGELHGHSGIARLALFYARAAACGLVDPSAGTGVANAG 251
QY 251 LVYFNGRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESTMIAPKVPDPESGELPAL 310
Db 252 LVYFNGRLAMSEDDLPYHVRVADGDGLTVGRYDFDGOGLGCAMIAHPKLDPATGELHAL 311
QY 311 SYDVSKPYLKYPRFSDGTSKSPDVEIQDPTMMDHFAITENFVVVPPQVVKLPEMI 370
Db 312 SYDVIKPYLKYFYFRPDGTSKSDVEIPLQPTMIHDFAITENFVVVPPHQVVKLQEML 371
QY 371 RGGSPVVDKKNVAKVFGIILDKYAESDSNIKWIDAPDCFCFHLNWAEEPEDEVVVGSC 430
Db 372 RGGSPVVDKKEKTSRFGVLPKHAADASEMAWVDVPCFCFHLNWAEDATGEVVVGSC 431
QY 431 MTPPDSIFNESDENLKSVLSEIRLNKLTGSESTRRPIISNEDQQVNLKAGMVRNMLGRKT 490
Db 432 MTPADSIFNESDERLSVLTEIRLDARTGSTRRAVLP-PSQENLEVGMVRNMLLGRES 490
QY 491 KFYALALAEPPKVPKVGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGGE-----EDEGY 545
Db 491 RYAYLAVAPWPKESFAKEDLSGTETKFEYGEGRFGGECFCFVMDPAAHAPRGEDDGY 550
QY 546 ILCFVHDEKTKWSELOIVANVLSLEVBATVKLPSRPYFGHGTFIGADDLAKQ 597
Db 551 VLTFFVHDERAGTSSELLVWNAADIRLEATVQLPSRVFPFGHGTFIGTQOELEAQ 602

RESULT 9
AAB72309
AC AAB72309 standard; Protein; 604 AA.
XX
XX AC AAB72309;
XX DT 16-MAY-2001 (first entry)
XX DE Neoxanthin cleavage enzyme-like protein amino acid sequence.
XX KW Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
XX NW NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
XX KW pathogen resistance; abscisic acid metabolism.
XX OS Zea mays.
XX FN W0200112801-A2.
XX PD 22-FEB-2001.
XX PF 17-AUG-2000; 2000WO-US22961.
XX PR 18-AUG-1999; 99US-0149656.
XX FR 23-MAY-2000; 2000US-0206405.
XX

```

(PION-) PIONEER HI-BRED INT INC.
(CURA-) CURAGEN CORP.

Bidney DL, Crasta OR, Hu X, Lu G;
WPI; 2001-211215/21.

Novel isolated defence-related signalling gene isolated from sunflower encoding neoxanthin cleavage enzyme, amino acid permease or glutamic acid-rich protein useful for increasing resistance of plant to a pathogen

Disclosure; Fig 1; 135pp; English.

This invention relates to defence-related signalling genes isolated from the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich protein (GRP). The signalling gene is useful for increasing the resistance of a plant to a pathogen such as fungus, virus, bacterium, nematode or insect (e.g. European corn borer), preferably *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a construct containing the gene into the genome of the plant. The gene is useful for regulating gene expression in a plant, in response to a stimulus such as infection with a pathogen, damage from a pathogen, hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid, oxalic acid or expression of a gene encoding oxalic acid oxidase. The genes are also useful for stem-preferred regulation of gene expression in a plant. The genes are useful in agriculture, particularly in the breeding of crop plants with improved agronomic traits, for modifying abscisic acid (ABA) metabolism and for modifying amino acid transport and content in plants. The present sequence represents a neoxanthin cleavage enzyme-like protein from *Zea* mays used in the characterisation of sunflower NCE.

Sequence 604 AA;

Query Match 61.3%; Score 1930; DB 22; Length 604;
Best Local Similarity 63.3%; Pred. No. 1.9e-186;
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSQSSDLSYCSLPMASRVTRKLNVSALHTP---PALHFPKQSSNSPAIVV---- 74
Db 17 PARSARAS-----NSVRFSPRAVSSVPPAECLQAPFHKPVADLPAPSRKPAIAVPGHA 71
QY 75 -KPKAKESNTKQNLKFORAAALDA-AEGFLVS-HEKHLPLKPTADPSVOIAGNAPVN 131
Db 72 AAPRKAEQKKQNLKFORAAALDAFEGFVANVLERPHGLPSTADPAVQIAGNAPVG 131
QY 132 EQVVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAKFEGHSA-SYA 190
Db 132 ERPPVHELPSGRIPFIDGVYARNGANPCFDPVAGHHLFDGDMVHALIRNGAAESYA 191
QY 191 CRFTQNRVFOERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDPAGTGVANAG 250
Db 192 CRFTETARLQERAIQRVFPKAIAGELHGHSGIARLALFYARAAACGLVDPSAGTGVANAG 251
QY 251 LVYFNGRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESTMIAPKVPDPESGELPAL 310
Db 252 LVYFNGRLAMSEDDLPYHVRVADGDGLTVGRYDFDGOGLGCAMIAHPKLDPATGELHAL 311
QY 311 SYDVSKPYLKYPRFSDGTSKSPDVEIQDPTMMDHFAITENFVVVPPQVVKLPEMI 370
Db 312 SYDVIKPYLKYFYFRPDGTSKSDVEIPLQPTMIHDFAITENFVVVPPHQVVKLQEML 371
QY 371 RGGSPVVDKKNVAKVFGIILDKYAESDSNIKWIDAPDCFCFHLNWAEEPEDEVVVGSC 430
Db 372 RGGSPVVDKKEKTSRFGVLPKHAADASEMAWVDVPCFCFHLNWAEDATGEVVVGSC 431
QY 431 MTPPDSIFNESDENLKSVLSEIRLNKLTGSESTRRPIISNEDQQVNLKAGMVRNMLGRKT 490
Db 432 MTPADSIFNESDERLSVLTEIRLDARTGSTRRAVLP-PSQENLEVGMVRNMLLGRES 490
QY 491 KFYALALAEPPKVPKVGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGGE-----EDEGY 545
Db 491 RYAYLAVAPWPKESFAKEDLSGTETKFEYGEGRFGGECFCFVMDPAAHAPRGEDDGY 550
QY 546 ILCFVHDEKTKWSELOIVANVLSLEVBATVKLPSRPYFGHGTFIGADDLAKQ 597
Db 551 VLTFFVHDERAGTSSELLVWNAADIRLEATVQLPSRVFPFGHGTFIGTQOELEAQ 602

```

Db 491 RYAYLVAEPWPESGFAKEDLSTGELTKEFYGEFGCEPVPMDPAAPRGGDOY 550
Qy 546 ILCFVHDEKTKWSELOIVNAVSLVEATVKLPSRVYFGHGTIGADDLAKQ 597
Db 551 VLTFVHDERAGTSELLVNAADIRLEATVQLPSRVFGRGTFTTQGLEAQ 602

```

RESULT 10

AAE04786 standard; Protein; 577 AA.

AC AAE04786;

DT 10-SEP-2001 (first entry)

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5.

KW Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;
stress tolerance; transgenic plant; plant breeding; antisense-therapy;
plant growth protectant.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX N-PSDB; AAD03398.

XX A DNA encoding a protein with a neoxanthin cleavage activity for
producing transgenic plants with improved or decreased stress tolerance

XX Claim 3; Fig 10; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their
corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
Neoxanthin cleavage enzyme is used for improving stress tolerance in a
plant when expressed in a plant cell. The invention also relates to
methods for increasing or decreasing stress tolerance in a plant by
introducing the DNA into the plant, and a transgenic plant into which a
neoxanthin cleavage enzyme is introduced. The improvement of stress
tolerance in plants is useful, for example in plant breeding. Neoxanthin
cleavage enzyme genes are useful for producing transgenic plants. An arid
land can be improved by growing transformant weed for several years and
then removing the weed by specifically lowering stress tolerance in the
weed by inducing an inducible promoter. The present sequence is
Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.
The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA
library using a cDNA of the CPD65 (CowPea Responsive to Dehydration)
gene isolated from cowpea plant as a probe.

XX Sequence 577 AA;

Query Match 52.8%; Score 1663.5; DB 22; Length 577;

Best Local Similarity 55.4%; Pred. No. 2.1e-159;

Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

Qy 24 LSSQSDLSYSCSLPMASRVTKLVSS-ALHTPPALHFPKQSSNSPAIVKPKAKESN'82

Db 9 LLPTKTSRSHLLPQKWNANISRRILNPFKIPKIPDLTSPVSP-----VKLKPTYPN 62

```

Qy 83 TKQNNLQRAAAALDAAE-GFLVSHKHLPLKTKTADPSVQIAGNFAPVNEQPVRRNL 141
Db 63 ---LNLQKLAATMLDKIESIVIPMEQNRPLPAPTDFAVQLSGNFAPVNECPVQNGLEV 119
Qy 142 VGKLPSDSIKGVYVANGANPLHEPVTGHHFFDGDGMHVK--FEHGSASACRFTQTNR 199
Db 120 VGQIPSCCLKGVYIRNGANPMFPPLAGHHLFDGDGMHVAIGFDN-QVSYSCTYKTNRL 178
Qy 200 VOERQLGRPYPPKAIHELHGTGIALMLFVARAAAGIVDPAGTGTVANAGLVYFNGRLL 259
Db 179 VQETALGRSVFPKPIGELHGHSGLARLALFTARAGIGLVDTGRGMGVANAGVVFNGRLL 238
Qy 260 AMSEDDLPYQVQITPENGDLTKVGRFDDGQLESTMIAPKPKVDPESEGFALSVDVWSKPY 319
Db 239 AMSEDDLPYQVKIDQGDLEITIGRFDDQIDSSVIAHPKVDTATGDLHLTSLYNLKKPH 298
Qy 320 LKYFRFPDGTGSPDVEIQDQPTMHDFAITENFVVVDDQVVKLPKEMIRGSPVYVD 379
Db 299 LRYLKFNTCGKKTRDVEITLPEPTMIHDAITENFVVIPOQMVFKLSEIRGSPVIYV 358
Qy 380 KMKVAFGILDKYAEDSNKWDAPDCFCFHLWNAMEEPETDE---VVVIGSCMTTPPD 435
Db 359 KEKVAFGVLSKQDLTGSDINWVDVDFCFHLWNAMEE-RTEEGDPVIVVIGSCMSPPD 417
Qy 436 SIFNESDENLKSVLSEIRNLKTGSTRRPIISNEDQOVNLEAGMVRNMLGRKTKAYL 495
Db 418 TIFSESGEPTRVLSSEIRLNNRTKESNRKVITG---VNLGAHINRSYVGRKSQFYI 473
Qy 496 ALAPFPKPVSGFAKVDLTGTVKXLYGDNRYGEGEPFLPGEGEDEDEGVTLCFVHDEKT 555
Db 474 AIADFPKCSGIAKVDIQNGTVSEFNYPSPFGGEPFVPEGEDEDEKGYVMGFVRDEK 533
Qy 556 WKSELOIVNAVSLVEATVKLPSRVYFGHGTFTIGADDLAKQV 598
Db 534 DESEFVVVDATDMKQVAARLPERVPYGFHGTFTVSENQLKEQV 576

```

RESULT 11

ABB92416

ID ABB92416 standard; Protein; 577 AA.

XX AC ABB92416;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1627.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
comprising aligning and comparing nucleic acid or amino acid sequences
from plant with nucleic acid or amino acid sequences from non-plant
organisms -

XX Claim 5; SEQ ID NO 1627; 261pp + Sequence Listing; English.

PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	99US-0157117.	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0158293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160980.
PR	26-JUL-1999;	99US-0145275.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145913.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145918.	PR	22-OCT-1999;	99US-0160989.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161404.
PR	28-JUL-1999;	99US-0146386.	PR	25-OCT-1999;	99US-0161405.
PR	02-AUG-1999;	99US-0146388.	PR	25-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0147038.	PR	26-OCT-1999;	99US-0161360.
PR	03-AUG-1999;	99US-0147038.	PR	26-OCT-1999;	99US-0161361.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161992.
PR	05-AUG-1999;	99US-0147192.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147260.	PR	29-OCT-1999;	99US-0162142.
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150684.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
Qy	42	SRVTRKLVSSALHTPPALHFPKQSSNSPAIVVKKPKAKESNTKQNLQFRAAAALDAAE	101	42	SRVTRKLVSSALHTPPALHFPKQSSNSPAIVVKKPKAKESNTKQNLQFRAAAALDAAE
Db	42	SPITNPSDNNDRNKPETH---NETNHTLVSSPKLRPEMTLALF---TTVEDVIN	94	42	SPITNPSDNNDRNKPETH---NETNHTLVSSPKLRPEMTLALF---TTVEDVIN
Qy	102	GFLVSHKULPLKPTADPSVQIAGNFAPVNEQPVERNLPVV-GKLPDSIKGVYVNGANP	160	102	GFLVSHKULPLKPTADPSVQIAGNFAPVNEQPVERNLPVV-GKLPDSIKGVYVNGANP
Db	95	TFIDP-----PSPSPVDPKHLSDNFAPVLDELPTDCEIHTGLTLPJSLNGYIINGNP	149	95	TFIDP-----PSPSPVDPKHLSDNFAPVLDELPTDCEIHTGLTLPJSLNGYIINGNP
Qy	161	LHEPVTGHHFFDGDGMVHAKPEHGSASVACRFTQTNRFQVQERQGRPVFPKAIQELHGH	220	161	LHEPVTGHHFFDGDGMVHAKPEHGSASVACRFTQTNRFQVQERQGRPVFPKAIQELHGH
Db	150	QFLPRGPFYHLFDGDMHAIKTHNGKATLCRSYVVTYKYNVEKQTGAPVMPNVFSGFNGV	209	150	QFLPRGPFYHLFDGDMHAIKTHNGKATLCRSYVVTYKYNVEKQTGAPVMPNVFSGFNGV
Qy	221	T-GIARMLFYARAAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVQITNGDLK	279	221	T-GIARMLFYARAAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVQITNGDLK
Db	210	TASVARGALTAARVLTGQYNPVNGIGLANTSIAFFSNELFALGESDLPYAVRLTSGDUE	269	210	TASVARGALTAARVLTGQYNPVNGIGLANTSIAFFSNELFALGESDLPYAVRLTSGDUE
Qy	280	TVGRFDFDGLSTMLAHPKVDPESEOLFALSYDVVSKYKLYKFFSPDGTSKSPVEI-Q	338	280	TVGRFDFDGLSTMLAHPKVDPESEOLFALSYDVVSKYKLYKFFSPDGTSKSPVEI-Q
Db	270	TIGRYDFDGLKAMSMTAHPKTDPIITGETAFAYGPV-PFPLYFRFDSAGKQROVPIFS	328	270	TIGRYDFDGLKAMSMTAHPKTDPIITGETAFAYGPV-PFPLYFRFDSAGKQROVPIFS
Qy	339	LDQPTWHDFAITENFVVVDQOVFK---LPEMIRGSGPVVDKKNVAFGILDKYABD	395	339	LDQPTWHDFAITENFVVVDQOVFK---LPEMIRGSGPVVDKKNVAFGILDKYABD
Db	329	MTSPSFLHDFAITKRAIFABIQLMRNMNLDLVLEGGSPVGTGDKTGRPGVIFRYAGD	388	329	MTSPSFLHDFAITKRAIFABIQLMRNMNLDLVLEGGSPVGTGDKTGRPGVIFRYAGD
Qy	396	SSNKKWIDAPDCFCFELWNWEEPEDEVVVTGSCMTPPDSIFNESDENKSVLSEIRLN	455	396	SSNKKWIDAPDCFCFELWNWEEPEDEVVVTGSCMTPPDSIFNESDENKSVLSEIRLN
Db	389	ESENKWFVPGFNIIHAINAWEDDNGSVLLAPNIMSIEHTLERND-LVHALVEKVKID	447	389	ESENKWFVPGFNIIHAINAWEDDNGSVLLAPNIMSIEHTLERND-LVHALVEKVKID

Query Match 30.7%; Score 968; DB 21; Length 595;

Best Local Similarity 36.7%; Pred. No. 1.1e-88; Indels 36; Gaps 13;

Matches 210; Conservative 106; Mismatches 22;

QY 456 LKGTSTRRPIISNEDQQVNLKAGWVRNMLGRKTKFAYLALAEPPKVGSGFAKVDLTG 515
 DB 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCRIVYAAIGDPMKISGVVKLDVSKG 502
 QY 516 E-----VKKHLVGNRYGGEPLFLPGGG-----BEDEGYILCFVHDEKTKWSELQIVNAV 566
 DB 503 DRDDCTVARRMYGCGYGGEPFFVARDPGNPEAEEDDGYVTVYVHDEVTSKFLVMDAK 562
 QY 567 S--LEVEATVKLPSRPVPGHGTIGADDLAK 596
 DB 563 SPELEIVAARVLRPRVPYGFHGLFVKESDLNK 594

RESULT 13
 ID AAE04783 standard; Protein; 595 AA.
 XX AAE04783;
 XX 10-SEP-2001 (first entry)
 DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2.
 KW Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
 KW plant growth protectant.
 OS Arabidopsis thaliana.
 XX EP1116794-A2.
 XX 18-JUL-2001.
 XX 11-JAN-2001; 2001EP-0300218.
 XX 13-JAN-2000; 2000JP-0010056.
 XX 11-JAN-2001; 2001JP-0003476.
 XX (RIKE) RIKEN KK.
 PA Iuchi S, Kobayashi M, Shinozaki K;
 PI WPI; 2001-400081/43.
 DR N-PSDB; AAD09395.
 XX A DNA encoding a protein with a neoxanthin cleavage activity for
 PT producing transgenic plants with improved or decreased stress tolerance
 PT -
 XX Example 10; Fig 10; 10pp; English.
 PS The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to
 CC methods for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant into which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present sequence is
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein.
 CC The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA
 CC library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration)
 CC gene isolated from cowpea plant as a probe.
 XX Sequence 595 AA;
 SQ

Query Match 30.7%; Score 968; DB 22; Length 595;
 Best Local Similarity 36.7%; Pred. No. 1.1e-88;

Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;
 QY 42 SVATKLVNSALHTPPALHPFKSSNSPAIVVFPKAKESNTKQWILFQRAAAALDAE 101
 DB 42 SPITNPSNNRRNKPKTLH---NRNHTLVSSPPKLRPEMTLALALF---TTVEDVIN 94
 QY 102 GLVLSHEKHLPLKPTADPSVQIAGNFAPVNEQPVRRNLPPVV-GKLPDSIKGYVVRNGANP 160
 DB 95 THIDP-----PSRPSVDPKHVLSONFAFVLDDELPTDCIEIHGTLPLSLNGAYIRNGNP 149
 QY 161 LHEPVTGHHFDDGDMHVAHFKHGSASACFTQTNRFVQERQLGRVFPFKAIGELHGH 220
 DB 150 QFLPRGPYHLFDGDMHAIKHNGKATLCRSRYKTYKYNVEKQTGAPWNPVFSGFNGV 209
 QY 221 T-GIARLMLFYARAAAGIVDPAHGTGVANAGLVFNGRLLAMSEDDLFYQVQITPNGDLK 279
 DB 210 TASVARGALTAARVLTGVNPNVNGIGLANTSLAFFSNELFALGESDLFAYVRLTESGDIE 269
 QY 280 TVGRFDFDQLESTMIAPKVDPESGELFALSVDVSKPYLYKFRFSPDGTGKSPDVEI-Q 338
 DB 270 TIGRYDFDGLKAMSMTAHPKTDPTITGETFAFYGVF-PPFLTYFRFDSAGKKQRDVPIS 328
 QY 339 LQOPTMMDHFAITENFVVVDQVVFK---LPEMIRGGSPVVYDKNKKVARFGILDKYAE 395
 DB 329 MTSPFLHDFAITKHAIFAETQLGRNMMLDLVLEGGSPVGTGDKTFLGVIEKYAGD 388
 QY 396 SSNIKWIDAPDCFCFLNANWEEPTDEVVIGSCWTPPDSIFNESDENLKSVLSEIRLN 455
 DB 389 ESEMKWFVEVGFNIIHAINAWDEDDGNSVLIAPNIMSGIETLERMD-LVHALVEKVKID 447
 QY 456 LKTGSTREPIISNEDQQVNLKAGWVRNMLGRKTKFAYLALAEPPKVGSGFAKVDLTG 515
 DB 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCRIVYAAIGDPMKISGVVKLDVSKG 502
 QY 516 E-----VKKHLVGNRYGGEPLFLPGGG-----BEDEGYILCFVHDEKTKWSELQIVNAV 566
 DB 503 DRDDCTVARRMYGCGYGGEPFFVARDPGNPEAEEDDGYVTVYVHDEVTSKFLVMDAK 562
 QY 567 S--LEVEATVKLPSRPVPGHGTIGADDLAK 596
 DB 563 SPELEIVAARVLRPRVPYGFHGLFVKESDLNK 594

RESULT 14
 ID AAB72306 standard; Protein; 595 AA.
 XX AAB72306;
 XX 16-MAY-2001 (first entry)
 DE Neoxanthin cleavage enzyme-like protein amino acid sequence.
 KW Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
 KW NCE; amino acid permease; AAP; glutamic acid rich protein; GAP;
 KW pathogen resistance; abscisic acid metabolism.
 OS Arabidopsis thaliana.
 XX WO200112801-A2.
 XX 22-FEB-2001.
 XX 17-AUG-2000; 2000WO-US22961.
 XX 18-AUG-1999; 99US-0149656.
 XX 23-MAY-2000; 2000US-0206405.
 XX (PION-) PIONEER HI-BRED INT INC.
 XX (CURA-) CURAGEN CORP.
 XX Bidney DL, Crasta OR, Hu X, Lu G;

DR WPI; 2001-211215/21.
XX Novel isolated defence-related signalling gene isolated from sunflower
PT encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
PT acid-rich protein useful for increasing resistance of plant to a
PT pathogen
XX
XX
XX Example 1; Fig 1; 135pp; English.
XX
XX This invention relates to defence-related signalling genes isolated from
CC the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
CC enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
CC protein (GRP). The signalling gene is useful for increasing the
CC resistance of a plant to a pathogen such as fungus, virus, bacterium,
CC nematode or insect (e.g. European corn borer). Preferably
CC *Sclerotinia spp.*, *Phoma spp.* or *Phymopsis spp.*, by stably incorporating a
CC construct containing the gene into the genome of the plant. The gene is
CC useful for regulating gene expression in a plant, in response to a
CC stimulus such as infection with a pathogen, damage from a pathogen,
CC hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
CC oxalic acid or expression of a gene encoding oxalic acid oxidase. The
CC genes are also useful for stem-preferred regulation of gene expression in
CC a plant. The genes are useful in agriculture, particularly in the
CC breeding of crop plants with improved agronomic traits, for modifying
CC abscisic acid (ABA) metabolism and for modifying amino acid transport and
CC content in plants. The present sequence represents a neoxanthin cleavage
CC enzyme-like protein from *Arabidopsis thaliana*, used in the
CC characterisation of sunflower NCE.
XX
XX Sequence 595 AA;
SQ
Query Match 30.7%; Score 968; DB 22; Length 595;
Best Local Similarity 36.7%; Pred. No. 1.1e-88;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;
QY 42 SRVTKLVNSALHTPPALFPKQSSNSPALVVKPKAKESNTKQMLFQRAAAALDAAE 101
DB 42 SPITNPQNDNRNKPXTLH---NRTHNLVSPFKLPKRPENTLALF-----TTVEDVIN 94
QY 102 GFLVSEKHLPLKPTADPSVQIAGNAPVNEQPVRRNLPPV-GKLDPDSKGVTVRNGANP 160
DB 95 TFDIP-----PSRPSVDPKRVLSNAPVLDDELPTDCEIIHGTLPLSLNGAYIRNGNP 149
QY 161 LHEBVTGHHFDDGGMVHAKFEHGSASYACRFTQTNRFVQERQLGRPVPEKALGELGH 220
DB 150 QFLPRGPVHLDGGMHLAIKHNGKALCSRYVTKYNNVEKQTGAPVNPVPSFGNGV 209
QY 221 T-GIARLMLFYARAAAGIVDPAGHTGVANAGLVYFNGRLAMSEDDLPHYQVQITPNDGLK 279
DB 210 TASVARGALTAARVLTGOYNPNVINGLANTSLAFPSNRLPALGSDLPYAVRLTESGDIE 269
QY 280 TVGRFPDQGLESTMIAPKVPDSGELFALSVDVSKPVLYKVERPSDGTGSPDYEL-Q 338
DB 270 TIGRYDFDGKLANSTAHKPTDPTITGETFAFRYGFV-PPFLTYPRFDSAGKKORDVPIFS 328
QY 339 LDQTMHDFAITENFVVVPDQVVFVK---LPENIRGSGSPVVDKKNVAFGLDKYAE 395
DB 329 MTSFSLDFAITKEHAIFAIIQLGMRNMLDLVLEGGSPVGTDNKTPRLGVIPKYAG 388
QY 396 SSNIKWDAPDPCFHLWNAWEETDEVVVIGSCMTFPPDSIFNESDENLKSVLSERLN 455
DB 389 ESEKMKFVPGFNIHAINAVDEDDGNSVLLIAPNIMSIEHTLERMD-LVHALVEKVKID 447
QY 456 LKTGEBSTRPTIISNEDQWNLGAGVNRNMLGRKTKFAYLALAPFPKPVSGFAKVDLTG 515
DB 448 LVTGIVREHPIASR-----NLDFAVINPAFLGRCSRYVYAIQDPMFKISGVVKLDVSKG 502
QY 516 E-----VKKHLYGNRNYGGEPLFIPGEGG----BEDGYILCFVHDEKTKWSELOINAV 566
DB 503 DRDDCTVARMYGGSCYGGEGFFVARDPFGNPEAEDDGYVVVYVHDEVGTGESKFLYMDAK 562
QY 567 S--LSEVATVKLPSSRVPGTGTGADDLAK 596

Db 563 SPELEIVAAVRLPRRVYPYGFHGLFVKESDLNK 594
RESULT 15
AAG31334
ID AAG31334 standard; Protein; 517 AA.
XX AAG31334;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37613.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127452.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.

```
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145125.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.4%; Score 959; DB 21; Length 517;
Best Local Similarity 39.2%; Pred.No.6.8e-88;
Matches 197; Conservative 98; Mismatches 183; Indels 24; Gaps 10;

QY 112 PLPKTADPSVQTAGNAPVNEQSVRENLFVV-GKLPDSIKGVYVRNGANPLHEPVTGHHF 170
Db 22 PSRPSVDPKHVSDNFPVLDLPPDCEIIHGTLPLSNGAYIRNGPNQFLPRGPYHL 81
QY 171 FDGDGMVHAVKFEHGSASYACRFTQTNRFVQRQLGRPVFPKAIKELHGT-GIARLMLF 229
Db 82 FDGDGMHLAIKHNGKATILCSRYVKTYKYNVEKQTGAPVMPNVFSFGNGVTASVARGALT 141
QY 230 YABAAAGIVDPAHGTGVANAGLVYFNGRLLANSEDDLPYQVQITPNGDLKTVGRDFDQG 289
```

